

**UNIVERSIDADE DE LISBOA**  
**Faculdade de Ciências**  
**Departamento de Informática**



**ENRIQUECIMENTO DE LIVROS FALADOS**  
**DIGITAIS COM IMAGENS**

**Hugo Miguel Ivo Braz Simões**

**DISSERTAÇÃO**

**MESTRADO EM INFORMÁTICA**

**2012**



**UNIVERSIDADE DE LISBOA**  
**Faculdade de Ciências**  
**Departamento de Informática**



**ENRIQUECIMENTO DE LIVROS FALADOS**  
**DIGITAIS COM IMAGENS**

**Hugo Miguel Ivo Braz Simões**

**DISSERTAÇÃO**

Projecto orientado pelo Prof. Doutor Luís Manuel Pinto da Rocha Afonso Carriço

**MESTRADO EM INFORMÁTICA**

**2012**





## **Agradecimentos**

Quero dedicar este trabalho à minha mãe, ao meu pai, ao meu irmão, ao Dudu, à avó Lena, aos meus amigos, e à minha maravilhosa Rituxa. Agradeço ainda à excelente escola que é a FCUL, a todos os colegas e amigos do HCIM/LaSIGE, e em particular, ao Luís Carriço por me ter ajudado a crescer pessoal e profissionalmente.



*Aos que cá estão. Aos que já cá não estão. E sobretudo, aos que virão.*



## Resumo

O objectivo deste trabalho é a concretização de uma ferramenta de anotação e classificação semântica e sintática de imagens, que assegure o enriquecimento da bancada de produção de livros falados digitais, com imagens.

É apresentada uma arquitectura de classificação e gestão de conteúdos multimédia genérica, capaz de lidar com o processamento de conteúdos multimédia, independentemente do seu tipo de conteúdos. No entanto, como o foco deste trabalho é a classificação de conteúdos com suporte em imagens, foi também, apresentada uma especialização desta arquitectura, para lidar exclusivamente, com elementos multimédia do tipo imagem.

A integração da ferramenta proposta com um repositório multimédia de imagens semanticamente indexadas, a par dos procedimentos de categorização de imagens, vão permitir o suporte à bancada de produção de livros falados digitais enriquecidos.

O resultado da aplicação da ferramenta proposta no seio da construção de livros falados digitais foi demonstrado através de um caso de estudo que revelou a importância, de se produzir uma ontologia de imagens que represente a informação processada, nas actividades de anotação e classificação de imagens, e na composição de imagens com texto.

Este trabalho foi suportado pela FCT - Fundação para a Ciência e Tecnologia, através do Laboratório de Investigação LaSIGE (POSI/ISFL/13/408) e do projecto RiCoBA. É fruto do trabalho referenciado nas linhas de investigação desenvolvidas no grupo HCIM/-LaSIGE, ao longo de dois projectos de investigação: numa primeira fase, pelo projecto *IPSOM* (POSI/34252/99) e, à posteriori, pelo *RiCoBA* (POSC/EIA/57038/2004).

**Palavras-chave:** Anotação de imagem, Classificação semântica de imagem, Ontologias de imagens, Livros falados digitais.



## Abstract

The aim of this work is the creation of a tool for semantic and sintatic image annotation and classification, that ensures the creation of rich digital talking books with images.

It is presented an architecture for the classification and management of generic multimedia to handle the multimedia content processing, regardless of type of content. However, as the focus of this work is the content classification of images, was also presented, a specialization of the above mentioned architecture to deal, exclusively with images.

The integration of a multimedia repository of semantically indexed images, coupled with images categorization procedures will assist the production of media enriched books.

The results of the proposed tool within the production of media enriched books was demonstrated through a case study that revealed the importance of producing a image ontology representing the information, for the annotation and classification activities as well as for the image composition with text activities.

This work was supported by FCT - Fundação para a Ciência e Tecnologia through LaSIGE research lab (POSI/ISFL/13/408) and *RiCoBA* project (POSC/EIA/57038/2004). It is the result of the work done in the research lines developed in the HCIM/LaSIGE group over two research projects: *IPSOM* (POSI/34252/99) and *RiCoBA*.

**Keywords:** Image annotation, Semantic image classification, Image ontologies, Digital Talking Books.





# Conteúdo

<b>Lista de Figuras</b>	<b>xvi</b>
<b>Lista de Tabelas</b>	<b>xix</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objectivos . . . . .	1
1.2 Contribuições e Resultados . . . . .	2
1.3 Estrutura da Dissertação . . . . .	3
<b>2 Panorâmica</b>	<b>5</b>
2.1 Contexto . . . . .	5
2.1.1 LFD e LFD Enriquecidos . . . . .	5
2.1.2 Reprodução de LFDE . . . . .	6
2.1.3 Produção de LFDE . . . . .	8
2.2 Ferramentas . . . . .	10
2.2.1 Photostuff . . . . .	15
2.2.2 AKTiveMedia . . . . .	17
2.2.3 M-OntoMat-Annotizer . . . . .	19
2.2.4 Comparação de ferramentas . . . . .	22
2.3 Sumário . . . . .	25
<b>3 Desenho e Concretização</b>	<b>27</b>
3.1 Arquitectura genérica de PCM . . . . .	27
3.1.1 Classificadores multimédia . . . . .	28
3.1.2 Compositores multimédia . . . . .	30
3.1.3 Gestor de conteúdos multimédia . . . . .	30
3.2 Casos de Uso . . . . .	31
3.3 Desenho do protótipo . . . . .	34
3.4 Unidade de processamento de imagem . . . . .	39
3.4.1 Extracção de características . . . . .	40
3.4.2 Ontologias multimédia . . . . .	44
3.4.3 Composição . . . . .	48

3.4.4	Gestão de conteúdos multimédia . . . . .	51
3.5	Sumário . . . . .	55
<b>4</b>	<b>Caso de Estudo</b>	<b>57</b>
4.1	Requisitos . . . . .	58
4.2	Produção e reprodução de LFDE . . . . .	59
<b>5</b>	<b>Conclusão</b>	<b>71</b>
5.1	Trabalho Futuro . . . . .	71
	<b>Bibliografia</b>	<b>79</b>
	<b>Índice</b>	<b>80</b>





# Lista de Figuras

2.1	Ferramenta de leitura de LFDE . . . . .	7
2.2	Arquitectura de produção de Livros Falados Digitais Enriquecidos . . . . .	9
2.3	Anotação e classificação de imagens no Photostuff . . . . .	17
2.4	Anotação e classificação multimédia no AKTiveMedia . . . . .	18
2.5	Anotação multimédia no M-OntoMat-Annotizer . . . . .	21
2.6	Extracção de descritores visuais no M-OntoMat-Annotizer . . . . .	22
3.1	Arquitectura genérica de PCM . . . . .	28
3.2	Caso de uso geral do protótipo . . . . .	32
3.3	Caso de uso anotação de imagens . . . . .	33
3.4	Vista da Ferramenta . . . . .	35
3.5	Interface com o utilizador . . . . .	36
3.6	Menu e barra de ferramentas . . . . .	36
3.7	Espaço de anotação e ferramentas automatizadas . . . . .	37
3.8	Espaço de classificação . . . . .	38
3.9	Espaço de composição . . . . .	39
3.10	Arquitectura da UPI . . . . .	40
3.11	Componente de extracção de características . . . . .	41
3.12	Anotação de imagens . . . . .	42
3.13	Desenho livre . . . . .	43
3.14	Operações de selecção e deslocamento e, de zoom . . . . .	44
3.15	Componente de ontologias multimédia . . . . .	45
3.16	Classificação de imagens . . . . .	46
3.17	Componente de composição de imagem . . . . .	49
3.18	Composição de imagem com texto . . . . .	50
3.19	Componente de indexação . . . . .	52
3.20	Componente de pesquisa . . . . .	53
3.21	Pesquisa de imagens . . . . .	54
3.22	Pesquisa de anotação de imagens . . . . .	55
4.1	Artefactos para a produção de LFDE . . . . .	59
4.2	Descrição da sincronização de áudio com texto . . . . .	60

4.3	Repositório multimédia . . . . .	60
4.4	Conteúdos no repositório multimédia . . . . .	61
4.5	Descrição do XML de sincronização áudio com o texto . . . . .	62
4.6	Ficheiro TXT e RTF do texto completo . . . . .	63
4.7	Ficheiro LAB de sincronização de áudio com texto . . . . .	63
4.8	Ficheiro TOC de descrição da tabela de conteúdos . . . . .	64
4.9	Conteúdo do ficheiro <i>AsMeninasDeLaMancha.RBK</i> . . . . .	64
4.10	Imagens que constam na obra . . . . .	64
4.11	Anotação da imagem da Capa do livro . . . . .	65
4.12	Anotação da imagem da Folha de Rosto do livro . . . . .	65
4.13	Anotação da primeira imagem do Capítulo I . . . . .	66
4.14	Composição da imagem da Capa do livro com o texto . . . . .	66
4.15	Ficheiro de sincronização de imagem com texto . . . . .	67
4.16	Colecção de entrada para o RBP . . . . .	67
4.17	Colecção de imagens para o RBP . . . . .	68
4.18	RBP a reproduzir o título do LFDE . . . . .	68
4.19	RBP a reproduzir a folha de rosto do LFDE . . . . .	69
4.20	RBP a reproduzir o capítulo I do LFDE . . . . .	69







# Lista de Tabelas

2.1	Comparação entre ferramentas estudadas . . . . .	24
3.1	Grau de automatização de operações sobre imagem . . . . .	44



# Capítulo 1

## Introdução

Nesta dissertação apresento o trabalho desenvolvido, no contexto do projecto *RiCoBA* (POSC/EIA/61042/04, 2004), integrado no grupo HCIM/LaSIGE da Faculdade de Ciências da Universidade de Lisboa. O trabalho incide sobre o desenvolvimento de uma ferramenta de suporte à classificação de imagens, de forma a contribuir para o enriquecimento de Livros Falados Digitais (LFD) com imagens.

### 1.1 Objectivos

O trabalho que se propõe assenta num conjunto de actividades com o propósito de enriquecer LFD com conteúdos baseados em imagens. Nesse sentido, o seu objectivo principal será a obtenção de informação e a classificação de regiões delimitadas de imagens com diferentes graus de complexidade, de forma a poderem ser referenciadas no seio de Livros Falados Digitais Enriquecidos (LFDE) e indexadas e manipuladas por uma bancada de produção desses livros.

Assim, importa analisar e compreender o funcionamento de sistemas existentes para a classificação de conteúdos de imagens e, com base neste estudo, definir ou ajustar uma representação adequada. Em particular, devem ser consideradas ontologias que garantam uma descrição fiel dos aspectos analisados cobrindo não só a perspectiva semântica dos conteúdos, mas também as propriedades visuais dos mesmos.

Noutra dimensão, deverão ainda incorporar, a capacidade de classificação genérica, independente de um domínio específico de aplicação. Naturalmente, a organização e correlação entre ontologias terá que ser considerada, não só relativamente às que proporcionam classificação de imagens, mas também entre essas e as que permitem a classificação de outros média, incluindo o texto, dando assim suporte à sua integração em LFDE.

Uma vez definidas as formas de classificação de imagens, o projecto propõe-se à definição e concretização de um repositório com capacidade de indexar, quer as descrições estruturadas, quer as próprias imagens. Sobre esse repositório, providenciar-se-á numa primeira fase, um protótipo de gestão de conteúdos indexados, que permita uma indexação

flexível, assim como uma pesquisa eficiente baseada no conteúdo. A definição desse repositório integrar-se-á com a especificação e concretização de uma versão mais alargada, de suporte à bancada de produção dos LFDE, que abranja os diversos tipos de componentes ilegíveis para a composição de LFDE. Sendo certo que o trabalho aqui proposto se centrará na componente de imagem, a sua articulação com os restantes componentes a desenvolver no contexto do projecto *RiCoBA* terá sem dúvida um impacto fundamental.

Finalmente, e constituindo um dos focos principais do projecto proposto, pretende-se definir e concretizar um protótipo de uma ferramenta de suporte à classificação de imagens, articulada com o repositório atrás mencionado, e que permita aos utilizadores de uma forma simples anotar, classificar e seleccionar regiões de imagens que constituam conteúdos relevantes. A identificação de regiões e a própria classificação seguirá fundamentalmente uma aproximação interactiva, controlada pelos utilizadores, mas, sempre que possível assistida. Esse suporte poderá passar pela integração do repositório de gestão de conteúdos, como forma de alimentar o processo de construção de LFDE.

Sendo esta, fundamentalmente uma ferramenta centrada na anotação semiautomática de imagens, desenvolver-se-á sobre o protótipo final um caso de estudo, por forma a aferir o seu papel na construção de LFDE com suporte em imagens, no contexto do projecto *RiCoBA*.

## 1.2 Contribuições e Resultados

As seguintes publicações apresentam as soluções para a problemática associada à classificação e anotação semântica de imagens expostas nesta dissertação, no âmbito dos LFDE. No entanto, não cobrem nem retratam a prova de conceitos, os resultados desta dissertação, e portanto, os contributos desta dissertação no seu todo.

### *Publicações*

Rui Lopes, **Hugo Simões**, Carlos Duarte, Luís Carriço  
'Rich Digital Books for the Web', In International Conference on Web Information Systems and Technologies, WebIST2007, Barcelona, Spain, March 2007.

Rui Lopes, Carlos Duarte, **Hugo Simões**, and Luís Carriço  
'On the road to Rich Digital Books', In Conferência Nacional em Interação Pessoa-Máquina, Interação2006, Braga, Portugal, October 2006.

### *Resultados*

- Ferramenta de suporte à classificação semântica de imagens;

- Anotação semântica manual e semiautomática de imagens;
- Representação de anotações sob a forma de ontologias de imagens;
- Enriquecimento da bancada de produção de LFDE.

### 1.3 Estrutura da Dissertação

O Capítulo 1 visa apresentar o contexto do trabalho e os seus objectivos. O Capítulo 2 descreve a perspectiva onde o trabalho exposto se enquadra e relata o estado da arte de trabalhos relacionado com a temática da anotação e classificação semântica de imagens. Neste capítulo é ainda descrito, o problema e apresentada uma arquitectura que pretende resolver o problema enumerado, tendo sempre por base o projecto *RiCoBA*. O Capítulo 3 descreve o desenho e a concretização do protótipo SICA aplicado ao projecto *RiCoBA*, assim como, os resultados da aplicação da ferramenta de suporte à classificação, no processo de construção de um livro falado digital enriquecido. É também apresentado o caso de estudo desta dissertação, no Capítulo 4. Finalmente, para terminar esta dissertação, o Capítulo 5 apresenta as conclusões do trabalho aqui descrito, dando igualmente linhas de orientação para uma boa continuidade do trabalho, até aqui conseguido.



# Capítulo 2

## Panorâmica

Neste capítulo começa-se por descrever o que são os LFD e os LFDE, o objecto específico do tema desta dissertação. De seguida, dá-se ênfase à forma como estes livros são reproduzidos por uma ferramenta de leitura de livros falados - o *Rick Book Player (RBP)*. Depois de apresentados os LFDE e a forma de reprodução dos mesmos, passa-se a descrever o modo como se produzem. Neste tópico apresenta-se ainda uma bancada tipo que considera os vários aspectos da produção de LFDE, expondo os componentes mais importantes. Sendo o foco desta dissertação, a criação de uma ferramenta de anotação e classificação de imagens que permita que se construam LFDEs com imagens, são também descritas e comparadas três ferramentas de anotação já existentes na comunidade. É nomeadamente, dado enfoque às características próprias que cada ferramenta apresenta e são comparadas as funcionalidades e tecnologias que adoptam.

### 2.1 Contexto

#### 2.1.1 LFD e LFD Enriquecidos

Os Livros Falados Digitais (Daisy Consortium, 2006; ANSI/NISO, 2002) são livros em formato digital que permitem acompanhar a leitura do texto com a reprodução digital da narração do seu conteúdo. Foram pensados como um meio de facilitar o acesso da comunidade de invisuais e pessoas com deficiências visuais aos livros. No decorrer do projecto *IPSOM* (POSI/PLP/34252/99, 1999) foram desenvolvidas ferramentas de alinhamento (identificação do tempo em que as palavras surgem na narração) e geração de LFD, nomeadamente, uma bancada flexível de produção desses livros, o DiTaBBu (Lopes et al., 2006), capaz de gerar livros em diversos formatos, com apresentação e formas de interacção multimodal. Estes desenvolvimentos abriram, entre outras, perspectivas para o alargamento do espectro de leitores originalmente apontados pelos LFD; para a extensão da experiência de leitura tradicional a situações e atitudes cognitivas distintas; bem como, a outro nível, para a capacidade de combinar e reutilizar componentes de livros em formas alternativas e digitais de publicação. Foram, assim, determinados os primeiros passos no

sentido de permitir o enriquecimento em diversas direcções de LFD e de facultar a criação de novas formas de contar histórias.

Os LFD enriquecidos estendem o conceito corrente de LFD introduzindo novas potencialidades como a apresentação e reprodução de novas formas de média em conjunto com a narração do livro. A criação do projecto *RiCoBA*, que surgiu como uma evolução natural do trabalho desenvolvido no âmbito do projecto *IPSOM*, pretende retirar a exclusividade do foco na parte textual do livro (com a respectiva narração), integrando a diversos níveis de importância outros média, como por exemplo imagens, vídeos ou som. Estes poderão complementar a informação textual, narrada ou não, ou substituí-la por completo (p. ex., livros de imagens) sendo, neste caso, a narração ou o texto o elemento secundário do livro. Em qualquer dos casos, os vários elementos nos diversos média deverão poder ser referenciados, na globalidade e nas suas partes (p. ex., uma palavra, um compasso, uma região de uma imagem), tal como tipicamente é feito nos LFD relativamente ao texto e ao áudio da narração. De realçar que, no contexto desta dissertação, a atenção estará centrada no enriquecimento de LFD com imagens.

### 2.1.2 Reprodução de LFDE

A flexibilidade do processo de produção de LFDs resulta em livros digitais personalizados para diferentes tipos de ferramentas de leitura de LFDs. Isto significa que são produzidos LFDs disponíveis para apresentação imediata em diferentes plataformas, ou que o resultado final do processo de produção pode ser fornecido a um artefacto de reprodução que trate da interacção e da apresentação.

No primeiro caso, a produção resulta em livros *Ready-to-Read* (Prontos-a-Ler) em plataformas de apresentação disponíveis facilmente no mercado. O formato de saída disponível, actualmente, na plataforma de produção inclui um formato de impressão e, outros que combinam texto e áudio, como o *SMIL* (Jeff Ayars, Dick Bulterman, *et al.*, 2001b) que pode ser executado no *RealPlayer*, ou o *HTML+TIME* (Schmitz et al., 1998) que está disponível para ser executado no *Internet Explorer*). Formatos em *Braille* para leitores invisuais são outros exemplos de formatos específicos, para leitores com necessidades especiais. No entanto, a possibilidade de apresentar o mesmo conteúdo com diferentes média e a operacionalização por leitores com necessidades visuais especiais, exigem capacidades de interacção multimodal. As diversas capacidades de personalização que devem estar disponíveis e a necessidade de poder gerir simultaneamente a transmissão de vários formatos de média, exige também a introdução de capacidades de adaptação.

Para ir de encontro a estes requisitos, foi desenvolvida uma ferramenta de leitura de LFDE com capacidades adaptativas multimodais, assente na bancada de trabalho *FAME* (Duarte and Carrico, 2006). Nesta ferramenta de leitura são identificados quatro componentes: conteúdo principal do livro, tabela de conteúdos (*TdC*), anotações e média de



apoio. Para cada componente é descrita a forma como cada um se interliga entre si, durante a reprodução dos livros. É também determinada a forma como cada componente é apresentado e a forma de interacção com o leitor. A ferramenta de leitura desenvolvida (ver Figura 2.1) promove uma melhor percepção do leitor em relação ao que está a ser narrado: um marcador realça o excerto que está a ser narrado; a secção que está a ser narrada ou o número de capítulo são realçados na *TdC*; a posição actual de leitura é definida por uma barra de deslocamento do conteúdo principal. O número do capítulo ou a presente secção podem ser narrados em qualquer momento, servindo de orientação aos leitores. Para perceber a estrutura do livro, o leitor pode navegar e examinar cuidadosamente a *TdC* a qualquer altura sem que a narração seja interrompida.

O suporte de anotações oferece ao leitor a possibilidade de criar e gerir notas e marcas

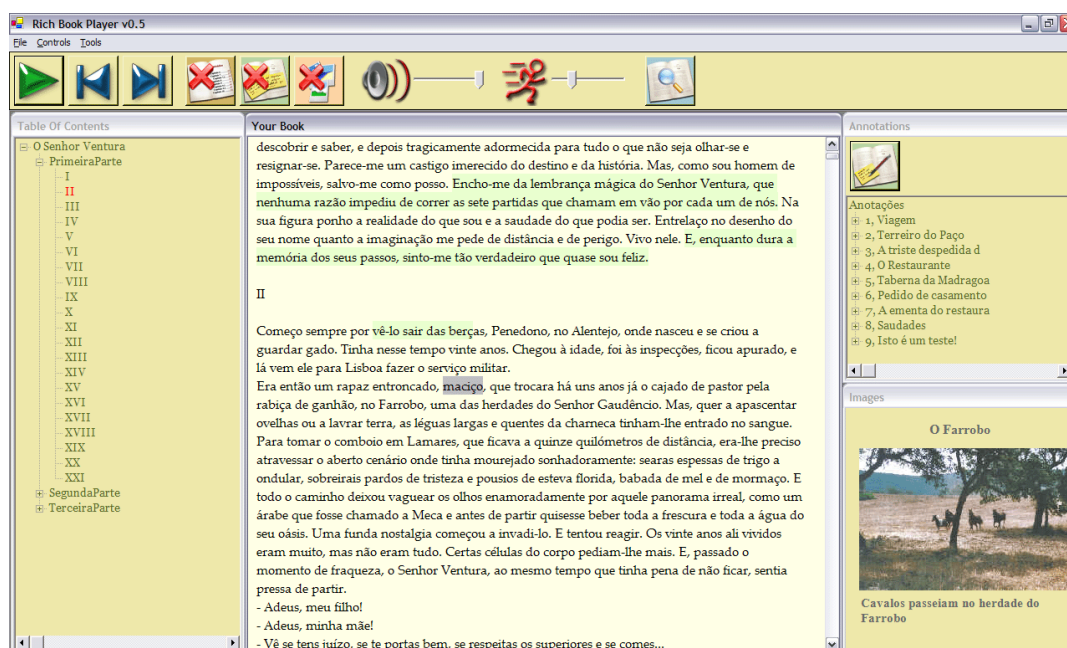


Figura 2.1: Ferramenta de leitura de LFDE

no livro. A percepção do leitor para com as anotações já realizadas, durante a narração do livro, é conseguida de duas formas: se a área de anotações está visível, o texto que tiver sido anotado é realçado no conteúdo principal do livro. Se pelo contrário, a área de anotação está invisível, o botão que permite mostrar ou esconder esta área, é realçado sempre que a anotação chega a um ponto do texto que tenha sido anotado.

No caso da janela de anotações estar activa e visível, o bloco de texto anotado é realçado com uma cor diferenciadora na área de texto; caso, a janela de anotações não esteja visível, é o botão que permite tornar visível, ou não esta janela, que é alvo de relevo e contraste, sempre que um bloco de texto anotado está a ser narrado. As componentes de média de apoio são tratadas de forma semelhante.

O uso de alertas áudio é também disponibilizado a quando da presença de anotações

ou imagens. Para minimizar as intervenções do leitor, a ferramenta de leitura pode adaptar o seu comportamento tendo em conta a apresentação de anotações e de outro material de suporte. Se ao alertar o leitor para a presença de imagens e anotações é usualmente encarado com um *feedback* positivo, então a ferramenta de leitura irá adaptar o seu comportamento para mostrar a imagem ou anotação, em vez de apenas, alertar o leitor para a sua presença. Mas, se o leitor decidir ignorar os alertas, haverá uma adaptação de comportamento, no sentido, da paragem dos mesmos.

### 2.1.3 Produção de LFDE

A produção automática e flexível de LFDE tem como base a arquitectura que está descrita na Figura 2.2 apoiada por um repositório multimédia, e baseada em documentos especificados ou transformados em linguagem *XML* (Sperberg-McQueen et al., 2008). Esta arquitectura é uma simplificação da apresentada no trabalho de Lopes et al. (2007), visto o foco deste trabalho incidir essencialmente nos aspectos de processamentos de conteúdos, em detrimento daqueles que se prendem com as dimensões comportamentais dos componentes de interacção e apresentação.

O processo de produção inicia-se, a partir de uma série de conteúdos que servem de matéria-prima, para a produção de um livro. Estes conteúdos podem ser da natureza daqueles que encontramos num qualquer livro, como é o caso, do texto e das imagens. No entanto, na construção de LFDE podem ser incluídos também, elementos baseados em áudio e mesmo vídeo. Uma vez armazenados em formato digital num repositório multimédia, estes conteúdos ficam disponíveis para ser objecto de transformação. O seu conteúdo pode ser modificado, simplificado e/ou aumentado, com o intuito de construir novos LFD mais ricos. O processo de produção culmina na entrega de um LFDE, adaptado às preferências do leitor, quer ao nível da interacção, quer ao nível da apresentação e que pode, por isso, ser reproduzido consoante o seu perfil.

#### Processamento de Conteúdos

O primeiro desafio da arquitectura de produção de LFDE passa pelo processamento de conteúdos que podem integrar um livro. Como estes conteúdos podem ser complexos e apresentarem um número variado de formatos, num primeiro momento são necessárias realizar operações de normalização. Estas operações podem ser tão simples como a conversão de texto de um PDF ou de HTML, num formato de texto simples integrado no formato de conteúdos específico da arquitectura, baseado nos modelos de hipermédia de referência de Hardman et al. (1994). No entanto, dependendo da riqueza dos conteúdos, pode ainda ser necessário proceder-se a operações mais complexas, nomeadamente, as de marcação de metadados que incluam análises semânticas e sintácticas de conteúdos multimédia. É neste aspecto que se centra o trabalho aqui apresentado. Em particular,

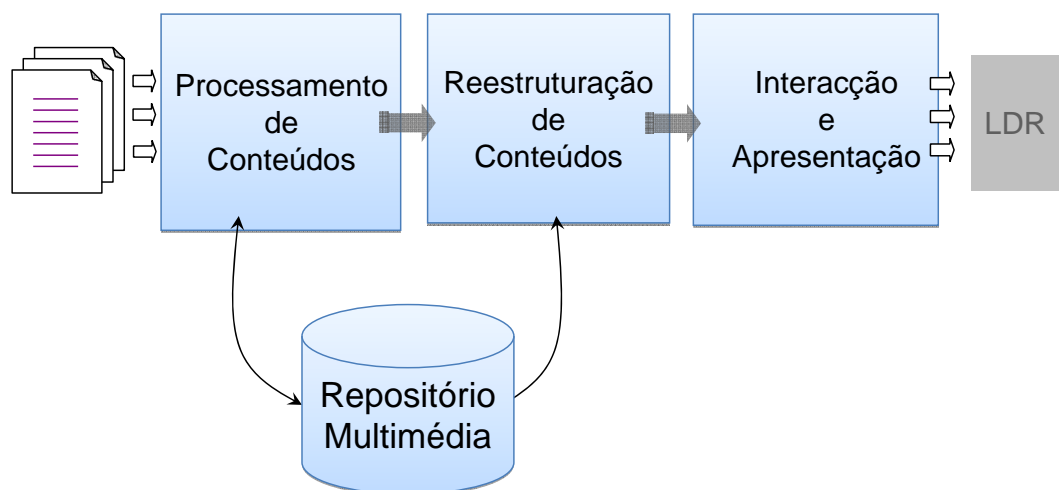


Figura 2.2: Arquitectura de produção de Livros Falados Digitais Enriquecidos

nas operações de anotação de conteúdos com suporte em imagens e da sua representação. Para dar suporte a estas operações, este componente encontra-se ligado a um repositório multimédia que através das suas capacidades de modelação e adaptação ricas, armazenam os elementos multimédia anotados e classificados visando facilitar a sua extracção e a integração na bancada de produção.

### Reestruturação de Conteúdos

O segundo desafio está relacionado com a reestruturação de conteúdos indexados, de modo a servir outros propósitos de LFDE. Devem ser fornecidos instrumentos que permitam aos diversos componentes da arquitectura, comunicar com o repositório multimédia. Aplicar-se operações de reutilização de conteúdos diversificados e, de enriquecimento do repositório com conteúdos multimédia, de livros já enriquecidos. Ou ainda ser aplicadas funções de extracção da estrutura, de conteúdos já normalizados. Permitindo assim, a criação de módulos estruturados independentes, como a extracção de uma lista de imagens.

### Formato de Saída

Outro desafio que se apresenta na arquitectura de produção, está relacionado com o formato de saída dos LFDE. Devem ser definidas funções para a conversão de estruturas de conteúdo normalizadas. Como devem ser tidos em conta vários cenários, devem ser suportados diferentes formatos (p. ex., *HTML+TIME*). Formatos mais ricos permitem a integração de capacidades de interacção e apresentação, enquanto outros, mais limitados, requerem apenas uma simples formatação do conteúdo. Assim devem ser aplicadas funções que possam ser aplicadas a um livro e, que não são mais do que construções personalizadas, para cada leitor (tais como, estruturas para marcação e anotação de conteúdos,

ou para aumentar e diminuir imagens). Paralelamente, devem ser tidas em conta, as diversas ferramentas de reprodução de LFDE e as suas capacidades de interacção e apresentação. Só assim, é possível fornecer ao leitor, um livro adequado ao seu perfil.

### **Interacção e Apresentação**

Depois da escolha da ferramenta de reprodução é necessário transformar o conteúdo do livro num formato de saída específico. Para isso, devem ser usados mecanismos para lidar com os aspectos relacionados com a interacção e a apresentação do conteúdo de um livro. A primeira preocupação relaciona-se com a configuração do formato de saída. Assim, devem existir funções que implementem mecanismos de interacção para entradas específicos, como o uso do rato, ou a capacidade de interacção por voz. A outra preocupação tem que vem com a forma, como um livro é apresentado. Aqui mais uma vez, têm de ser afinadas um conjunto de configurações, de modo, a manter a coerência da interface que se quer, com o leitor. Nomeadamente, aquelas que se referem com a definição de perfis e regras de apresentação, para diferentes formatos de saída.

O foco deste trabalho é a criação de uma ferramenta de anotação e classificação de imagem e sistemas de suporte. Portanto, de seguida, vão ser apresentadas três ferramentas bem conhecidas dentro desta área e da comunidade ligada à anotação e classificação de conteúdos multimédia em geral. Pretendendo-se desta maneira, comparar cada uma das ferramentas descritas, em relação às funcionalidades e às tecnologias, quer as que estão subjacentes à sua criação, quer aquelas intrinsecamente ligadas à criação dos artefactos, por estas produzidas.

## **2.2 Ferramentas**

No âmbito do IPSOM, e sobretudo do RiCoBa, foram já desenvolvidos protótipos para reprodução de LFDE. O RiPlay (Rodrigues et al., 2006), por exemplo, oferece novas formas de leitura destes livros. O trabalho de Duarte and Carrico (2005) permite a adaptação da forma de apresentação às características de uso, incluindo o aparecimento de imagens e a forma de o fazer. No entanto, os elementos de enriquecimento dos livros são escolhidos e retirados de diversas fontes dispersas e externas ao projecto, de uma forma *ad hoc*. Para além disso, são inseridos de forma pouco estruturada e manual nos conteúdos originais, durante as tarefas de reprodução do livro e, finalmente, as imagens são vistas como caixas negras sem qualquer tipo de acesso indexado ao seu conteúdo.

Uma solução típica para as limitações anteriores é a criação de um repositório de média (Subrahmanian, 1998). A sua integração quer na reprodução quer na produção de LFDE, pressupõe a reutilização de diversos tipos de média (Cybulski and Linden, 1999), o que requer uma boa gestão de conteúdos. Por isso, a informação deve estar

estruturada num formato normalizado. Devem também existir mecanismos flexíveis para a identificação, classificação e organização dos elementos multimédia reutilizáveis, que possam vir a integrar o repositório.

Tendo em conta os diversos tipos de média que serão objecto de indexação e a necessidade de classificação de elementos multimédia, é fundamental ter em consideração as abordagens centradas em mecanismos da *Web semântica* (Berners-Lee et al., 2001; W3C Semantic Web Activity, 2008). Nomeadamente, aqueles que estão relacionados com as anotações multimédia, e em particular, com as anotações de imagens (W3C SWBPD MM Task Force, 2006). Por forma a serem representadas as descrições visuais e semânticas dos elementos multimédia a indexar, devem ser utilizadas ontologias multimédia (Husemann and Vossen, 2004; Staab et al., 2001). E assim, obter-se uma representação estruturada e normalizada de conteúdos. A utilização de ontologias genéricas baseadas na estrutura do *MPEG-7 Visual Part* (Sikora, 2001; Simou et al., 2005) permite uma descrição mais exacta da informação. A par disto, o uso de uma linguagem formal como o OWL<sup>1</sup> para exprimir ontologias, possibilita uma uniformidade de descrições de conteúdos para um dado domínio.

São estas soluções apresentadas que ajudam à resolução da interoperabilidade semântica de anotações entre diferentes ferramentas de anotação, minimizam a distância semântica e evitam diferentes interpretações do conhecimento para um dado contexto. Para além disso, tornam menos complexo o processo de extracção de informação baseado em conteúdo. Assim, a criação de uma ferramenta de anotação de imagens que apoie a criação de LFDE e sua integração na bancada de produção de LFDE, deve tirar partido destas soluções.

Sendo certo que a existência de um repositório é fundamental para um processo estruturado e coerente de reprodução adaptativa ou produção de LFDE, também o é, como acontece já nos LFD, a criação do *corpus* de elementos que estarão na base dessa reprodução/produção. De facto, no âmbito dos projectos acima mencionados - *IPSOME* e *RiCoBA*, a bancada de reprodução prevê uma primeira fase de processamento de áudio e de texto que permitirá identificar esses elementos tornando-os elementos indexáveis. Do mesmo modo, ao introduzir outras formas de conteúdo é indispensável encontrar mecanismos que permitam a indexação dos outros tipos de média, permitindo assim a alimentação de um repositório rico. Naturalmente, será desejável a existência de processos tendencialmente automáticos ou semiautomáticos de classificação desses conteúdos por razões de escala, tão presente na imagem e no vídeo como nas narrações áudio originais dos LFD. Todavia, os desafios nestas áreas são grandes. As soluções existentes passam por uma análise sintáctica automática e por outra semântica assistida. A primeira requererá o uso de técnicas de processamento de imagem Gonzalez and Woods (2002); Jahne (2002); Pratt (2001) e visão por computador Shapiro and Stockman (2001); Forsyth and Ponce (2002) para extrair aspectos mais relevantes do conteúdo de uma determinada imagem

---

<sup>1</sup><http://www.w3.org/TR/owl2-overview>

que se pretende classificar. Este processamento terá de se basear em propriedades visuais, como cores, formas, texturas. A análise semântica prende-se com anotações baseadas no conteúdo semântico, ao invés de se basear apenas no conteúdo visual.

Nesta secção pretende-se avaliar diversas ferramentas de anotação multimédia, em relação a uma série de características que emergem do seu contexto de utilização. Para tal, realizaram-se estudos que tomaram em consideração as seguintes ferramentas de anotação, tanto manuais, como baseadas em ontologias: Photostuff, AKTiveMedia e M-OntoMat-Annotizer.

As ferramentas multimédia analisadas tanto servem diferentes propósitos, como lidam com diferentes tipos de conteúdos, tornando assim, difícil a formação de conclusões comuns sobre os princípios de *design* por detrás de ferramentas deste género, uma vez que é sobretudo o domínio projectado e o contexto de aplicação a que se destinam, que determina estes princípios. Enquanto o PhotoStuff e o AKTiveMedia endereçam aspectos essencialmente descritivos de anotações, o M-OntoMat-Annotizer é principalmente orientado à análise.

As ferramentas de anotação foram criadas com o intuito de produzir anotações de documentos. Mas, a grande maioria das ferramentas exigem uma anotação manual. No entanto, muito poucas pessoas vão anotar páginas Web à mão. Porque é uma tarefa exaustiva, entediante e morosa. Para além disso, é ineficiente, se a quantidade de documentos que se quiser anotar for, de facto, muito grande. É bastante controverso este tipo de anotação manual, já que, existe também o risco de se ter várias pessoas a anotar a mesma informação, de formas diferentes. Por exemplo, ao nível da interpretação semântica, da contextualização ou de outros aspectos, como os visuais, que cada um faz ao deparar-se com a necessidade de anotar qualquer coisa que seja. A discordância entre utilizadores pode resultar uma dispersão de dados e tornar a extracção de informação bem mais difícil e ambígua. Ainda mais se, se tiver uma bancada de produção de conteúdos e se, se pretender utilizar elementos multimédia anotados que contenham determinadas características, em particular.

Portanto, é de importância extrema a criação de processos que de alguma forma automatizem os processos de anotação, de armazenamento e de extracção. Por um lado, pretende-se que ajude na anotação manual, por outro, que substitua a necessidade de ter humanos a anotar conteúdos multimédia.

O processo de anotação de texto tem em conta a extracção de entidades, campos, relações e eventos presentes num texto. Enquanto o processo de anotação de imagens é focado na semântica dos conteúdos apresentados na imagem e na sua análise sintáctica. Em ambos os processos pode recorrer-se ao uso de ontologias, quer para nos processos de anotação quer na extracção de instâncias de anotações, ou ainda, na classificação e

agrupamento de elementos multimédia. Por exemplo, no agrupamento de imagens de acordo com as disposições visuais dominantes.

No que respeita aos tipos de anotação multimédia, estas podem ser divididas em 3 categorias:

- Anotação livre - é muito expressiva e vista como uma anotação ordinária para o humano, no entanto, oferece menos semântica formal e, portanto, o seu processamento por um computador torna-se muito mais difícil;
- Etiquetagem - as etiquetas ou marcadores não têm semântica formal, mas podem ser úteis se, se usar um vocabulário fixo para as anotações multimédia;
- Ontologias - as que fornecem sintaxe e semântica para a definição de vocabulários de domínios complexos; possibilitam a inferência de conhecimento adicional; pode ser vista como uma solução para os problemas de interoperabilidade entre sistemas.

A criação de bibliotecas digitais com conteúdo inteligente só é possível, através da criação de ontologias para o conteúdo digital arquivado nestas bibliotecas. As ontologias criadas e sua anotação semântica permitirão pesquisar todo o tipo de materiais usando técnicas da Web semântica.

As ontologias são semanticamente mais ricas do que vocabulários controlados, léxicos, glossários e taxonomias porque podem ser usadas para expressar uma quantidade diversificada de relações entre conceitos. Além disso, como as ontologias podem expressar essa riqueza semântica de uma maneira formal, são pois mais adequadas na geração de conhecimento para interpretação e leitura por uma máquina.

Para além disso, apesar do foco principal desta secção estar na análise descritiva de cada uma das ferramentas em avaliação, em relação às suas capacidades de anotação de imagens, é também dado a conhecer, não de um modo tão exaustivo, é certo, as características destas ferramentas, em relação às suas capacidades de anotação de texto. No caso da anotação de texto, o objectivo principal é o reconhecimento da semântica do texto anotado em análise (ou de partes de texto, como sejam, parágrafos, frases, palavras, etc.). Por outro lado, as anotações multimédia, onde se incluem as anotações de imagens, têm como pressuposto a definição manual da semântica dos respectivos conteúdos anotados.

No caso das ferramentas de anotação multimédia, o formato da ontologia de entrada não é necessariamente o mesmo que o formato dos metadados produzidos. Esta é uma consequência directa do processo de anotação de elementos multimédia, pelo menos, tal como é concretizado pelas ferramentas que se encontram disponíveis nos dias que correm. A ontologia de entrada serve principalmente como vocabulário controlado para instanciar conceitos e relações presentes no conteúdo anotado. Em termos práticos, isso significa que as ferramentas que usam uma ontologia de entrada, por definição, não podem ir para além da expressividade que a ontologia carregada reflecte, podendo apenas

produzir-se simples asserções de conceitos e relações. Não permitindo a materialização destas asserções, a concretização de abstrações de conceitos nem a criação de restrições de propriedades de uma relação. Por este motivo, a maioria deste tipo de ferramentas utiliza como formato de saída dos seus metadados, uma especificação em RDF, e assim, separa estes aspectos, da linguagem ontológica fornecida pela ontologia de entrada.

Todas as ferramentas analisadas utilizam diversos formatos padrão, como XML<sup>2</sup>, RDF<sup>3</sup> e OWL para guardar anotações. Para além destes, Handschuh and Staab (2002) e Klien (2007) adoptam outros tipos de formatos padrão para descrever os metadados, como por exemplo, Dublin Core (Hanisch, 2005), VRA Core (VRA, 2002) e ISO 19115 (ISO, 2008), aumentando assim a probabilidade de sucesso do conteúdo anotado ser encontrado. No entanto, quando estão em causa pesquisas semânticas ou baseadas na semântica das anotações produzidas, é imprescindível que as anotações estejam guardadas em RDF e OWL.

A característica comum partilhada entre as ferramentas multimédia de anotação examinadas, é o facto de elas tratarem de anotações manuais. A maioria das ferramentas de anotação multimédia existentes proporciona apenas facilidades de anotação manual. Contudo, ao longo da última década a comunidade científica ligada à anotação multimédia baseada em conteúdo, centrou todos os seus esforços na automatização ou na criação semiautomática de anotações multimédia, levando a uma multiplicidade de iniciativas e esforços de investigação, incluindo uma máquina de aprendizagem e metodologias de conhecimento assistido. No entanto, ultrapassar a chamada distância semântica, apresenta ainda grandes desafios, especialmente ao considerarmos o domínio e a independência da aplicação. Exemplos destes desafios são agora enumerados:

- A anotação multimédia cruzada entre diversos tipos de elementos multimédia;
- A visualização de ontologias;
- A geração de metadados baseados em ontologias, susceptíveis de armazenamento, extracção e reutilização;
- A pesquisa de metadados;
- O processo de anotação assistido;
- As anotações de regiões baseadas no conteúdo de imagens;
- A extracção dos metadados existentes que estão embutidos nas próprias imagens e serialização dessa informação nos metadados gerados no final do processo de anotação;

---

<sup>2</sup><http://www.w3.org/TR/REC-xml>

<sup>3</sup><http://www.w3.org/TR/rdf-syntax-grammar/>



- A geração automática de metadados;
- A reutilização de conhecimento e inferência de conhecimento baseado no conhecimento existente;
- Mecanismos de partilha de conteúdo.

Existe um conjunto alargado de ferramentas que têm sido desenvolvidas, umas comerciais e outras em ambientes de investigação académica, com o objectivo de fornecer funcionalidades que permitam a criação de metadados semânticos para conteúdos na Web, como por exemplo, o Annotea (Kahan et al., 2001) e o SHOE (Heflin and Hendler, 2000). Muitas ferramentas proporcionam um ambiente de anotação apenas para texto simples de páginas Web. No entanto, há a necessidade de anotar outros tipos de conteúdos disponíveis na Internet.

### 2.2.1 Photostuff

A motivação do PhotoStuff (Halaschek-Wiener et al., 2005) é a de uma ferramenta que proporcione aos utilizadores a possibilidade de efectuar anotações sobre determinadas regiões numa imagem, tendo como base o uso de uma ontologia OWL<sup>4</sup> genérica. Que tenha ainda, a capacidade de publicar os metadados gerados das regiões anotadas de forma automática, e disponibilizá-los via Web. A anotação de imagens é independente da plataforma e a ontologia que é utilizada fornece a expressividade necessária para associar e descrever regiões de uma imagem, bem como outras informações sobre a imagem, como data de criação, o formato da imagem e outras propriedades próprias da mesma.

O PhotoStuff<sup>5</sup> que nasceu do projecto *Mindswap*<sup>6</sup> fornece uma arquitectura simples, baseada em *plugins* para que se possa estender as suas funcionalidades, por exemplo, adicionar suporte para um novo tipo de armazenamento de dados ou de elementos multimédia.

Dois destes *plugins* pretendem alargar o suporte de armazenamento de dados, para além de arquivos baseados em URIs e da tecnologia por trás do portal Mindswap. Há também *plugins* para lidar com fontes de dados de Sesame (Broekstra et al., 2002) e Kowari (Wood, 2005). Estes *plugins* fornecem suporte à utilização de qualquer um dos sistemas acima mencionados, ao nível do armazenamento, extracção e análise de metadados. No entanto, existe um outro *plugin* que assegura o suporte de linguagem natural à interface principal do PhotoStuff, oferecendo descrições de média e das suas representações em linguagem natural.

As classes da ontologia utilizada podem ser arrastadas para qualquer região, ou para a própria imagem, criando uma nova instância da classe seleccionada. Alternativamente, o

<sup>4</sup><http://www.mindswap.org/2005/owl/digital-media>

<sup>5</sup><http://www.mindswap.org/2003/PhotoStuff>

<sup>6</sup><http://www.mindswap.org>

utilizador pode especificar primeiro as regiões da imagem, e arrastar uma classe da árvore de classes para uma dessas regiões, de modo a criar anotações de regiões associadas a classes. Por outro lado, pode criar anotações usando as instâncias pré-existentes, arrastando um item da lista das instâncias, para examinar a imagem ou as suas regiões. É ainda possível, processar múltiplas anotações de uma só vez.

O PhotoStuff garante ainda a possibilidade de se extrair e codificar a informação de metadados, representados em Dublin Core e EXIF (JEITA, 2002) embutidos em arquivos de imagens, em formato RDF/XML (Beckett, 2004). Garante também que se recorra ao uso de marcadores para manter ligações com ontologias e elementos multimédia que são de interesse geral e que se deseje aceder de forma rápida.

Na Figura 2.3 pode ser observada a interface do Photostuff, que no geral, assenta nas seguintes características e funcionalidades:

- Anotar regiões de imagens recorrendo à utilização de ontologias personalizadas e predefinidas pelo utilizador;
- Usar uma variedade de ferramentas de desenho de regiões, para que os utilizadores possam ser capazes de destacar as regiões em torno de porções de imagens;
- Extrair metadados embutidos nas próprias imagens e serialização dessa informação nos metadados gerados de uma determinada anotação;
- Suportar o carregamento de ontologias e de imagens tanto, a partir da Web, como do disco local, enquanto as anotações produzidas podem ser armazenadas igualmente, em disco ou publicadas num portal;
- Carregar múltiplas ontologias de uma só vez, e assim, tirar proveito da marcação e associação de regiões de imagens com conceitos pertencentes a qualquer das ontologias suportadas e carregadas;
- Carregar instâncias de anotações a partir de qualquer URI na Web, podendo depois as suas representações referenciar outras instâncias existentes, possibilitando a reutilização de definições de metadados provenientes de outras fontes;
- Interagir com um portal de Web semântica permitindo a extracção de todas as instâncias de anotações submetidas nesse portal, assim como a colocação dos metadados gerados pelas anotações guardados em RDF/XML, e o carregamento das próprias imagens, por forma, a que as instâncias possam ser referenciadas por um URI;
- Visualizar e pesquisar imagens com base na sua semântica, após a submissão dos seus metadados no portal de Web semântica.

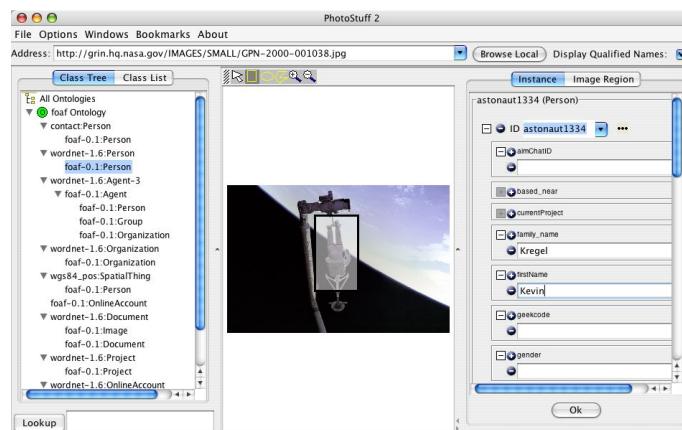


Figura 2.3: Anotação e classificação de imagens no Photostuff

### 2.2.2 AKTiveMedia

O AKTiveMedia (Chakravarthy, 2006) utiliza as tecnologias da Web semântica e de processamento de linguagem natural para ajudar os utilizadores a anotar conteúdo, com a vantagem de permitir anotar uma grande variedade de tipos de média diversificada, como por exemplo, texto e imagens. A forma de interacção dos utilizadores com a ferramenta é feita à através de anotações livres de texto e de imagens e, é baseada em ontologias.

O AKTiveMedia<sup>7</sup> é um sistema de anotação multimédia baseada em ontologias com o foco no utilizador. Tem por objectivo automatizar o processo de anotação, por exemplo, através da sugestão de factos subjacente a uma dada imagem, e de uma forma interactiva minimizando o esforço do utilizador aquando do processo de anotação de conteúdos. O funcionamento deste sistema é feito à custa da interacção com serviços Web e de pesquisas a um repositório central de anotações multimédia, conseguindo assim, dar respostas para um determinado contexto específico e, possibilitando ainda, a reutilização de anotações.

A par dos serviços Web, as tecnologias ligadas ao processamento de linguagem proporcionam uma forma de contextualizar anotações, inferindo sugestões, tanto a partir de ontologias como da reutilização de anotações, previamente feitas. Este mecanismo é muito importante para ajudar o utilizador, já que a ontologia que usa numa determinada anotação é alvo de uma filtragem, de modo, a apresentar apenas os conceitos de nível mais genérico. Uma das características interessantes do AKTiveMedia é o facto de permitir a indexação e a extracção semântica de informação gráfica. Para além disso, o AKTiveMedia pode funcionar como plataforma base de interacção com o utilizador, e por detrás, pode estar algo como por exemplo, o PhotoCopain (Tuffield et al., 2006).

Tal como a Figura 2.4 apresenta, as imagens no AKTiveMedia podem ser anotadas como um todo ou podem ser anotadas apenas partes ou regiões específicas que tenham

<sup>7</sup><http://www.aktors.org/technologies/aktivemedia>

maior relevância. Primeiro que tudo, deve ser atribuído para cada imagem ou parte desta, um título e uma descrição juntamente com comentários em texto livre, contextualizando-as de alguma forma, consoante o objectivo inerente à anotação em causa. Todos estes elementos que se adicionam às imagens anotadas são extraídos e vistos como metadados e, são também eles anotados, usando uma estratégia de representação de conhecimento que relaciona e combina texto com imagens.

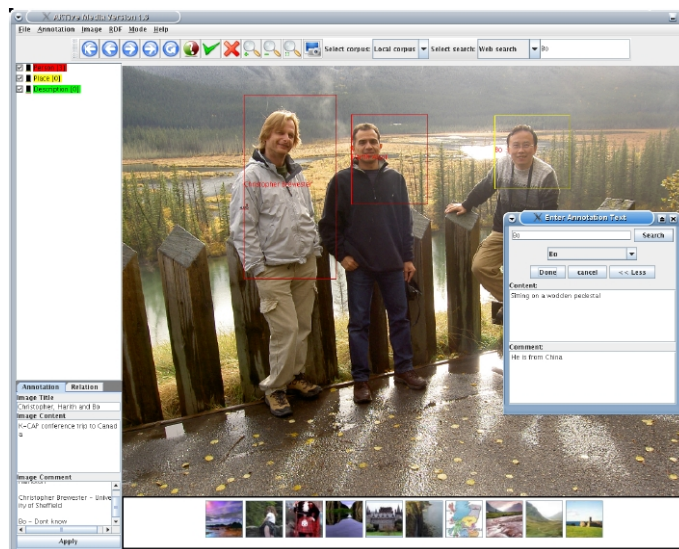


Figura 2.4: Anotação e classificação multimédia no AKTiveMedia

As anotações de partes específicas de imagens ou imagens completas são capturadas pelo sistema e descritas num único RDF. Isto significa que cada RDF gerado diz respeito, a uma associação entre um conceito em particular ou mesmo uma instância da ontologia usada na anotação, e uma determinada região da imagem ou mesmo a uma imagem inteira. Portanto, no caso desta ferramenta, o foco da representação do conhecimento passa por uma ontologia fixa desenvolvida em OWL (Horrocks et al., 2003) e pelo uso de RDF.

No que diz respeito, à pesquisa de conteúdos o AKTiveMedia usa a tecnologia SPARQL (Prud'hommeaux and Seaborne, 2007) em que a sugestão de conhecimento ao utilizador é feita através de mecanismos de busca inteligente, nomeadamente baseados nesta tecnologia. Este mecanismo de sugestão permite minimizar o esforço do utilizador quando tem que lidar com ontologias complexas, pois, ao utilizador é sugerido e apresentado uma ontologia mais genérica e simples, enquanto o próprio mecanismo trata da especialização dessa ontologia subjacente, de forma transparente para o utilizador. O SPARQL pode ser usado para expressar pesquisas em torno de diversas fontes de dados, independentemente de os dados estarem armazenados nativamente em RDF, ou vistos como RDF após terem sido processados por um qualquer tipo de processo intermédio.

De seguida, são apresentadas as principais funcionalidades que caracterizam a ferramenta AKTiveMedia:

- O suporte à anotação de regiões, de colecções inteiras de imagens e à anotação colaborativa de conteúdos;
- A anotação relacional de imagens, isto é, pode se especificar que uma parte X tem uma falha Y, quando se está anotando um motor Z;
- O suporte a anotações de texto/HTML;
- A anotação de partes, do todo ou de um conjunto de imagens baseada em ontologias existentes ou definidas pelo utilizador;
- A anotação entre diferentes tipos de média;
- A capacidade para importar diferentes formas de representação de ontologias, por exemplo, OWL, RDF, RDFS<sup>8</sup> ou DAML<sup>9</sup>;
- O suporte para diversos tipos de formatos de imagens;
- A representação semântica do vocabulário em DOAP RDF (Dumbill, 2004);
- A integração com serviços Web, na procura de imagens relevantes para o processo de anotação;
- A importação de RDFs e a facilidade de exportação de dados anotados em RDF, para acesso ou publicação dessas informações para o portal de Web semântica;
- A pesquisa de imagens com conteúdos equivalentes ou semelhantes, permitindo ao utilizador explorar as imagens pesquisadas sob a forma de um cronograma temporal.

### 2.2.3 M-OntoMat-Annotizer

O M-OntoMat-Annotizer (Petridis et al., 2006) é uma ferramenta de anotação multimédia desenvolvida no âmbito do projecto *aceMedia*<sup>10</sup>. Trata-se de uma extensão da plataforma CREAM de Handschuh and Staab (2002); Handschuh et al. (2002) e da sua implementação de referência, chamada OntoMat-Annotizer.

O projecto *aceMedia* visa criar uma plataforma, que combina os avanços no conhecimento, na semântica e nas tecnologias de processamento multimédia, de modo a garantir uma maior automatização da análise, da anotação, e da adaptação de conteúdos multimédia. Os objectivos deste projecto passam também por fornecer as ferramentas necessárias para auxiliar o utilizador na gestão de conteúdos digitais. Permite ao utilizador

---

<sup>8</sup><http://www.w3.org/TR/rdf-mt>

<sup>9</sup><http://www.daml.org>

<sup>10</sup>[http://cordis.europa.eu/ist/kct/acemedia\\_synopsis.htm](http://cordis.europa.eu/ist/kct/acemedia_synopsis.htm)

lidar com grandes volumes de informação, para que estes não só tenham acesso a conteúdo de diversos tipos, mas também a ferramentas que permitam criar, por eles próprios, diferentes tipos de conteúdos. Facilitando assim, o processo de procura de conteúdo, a criação de novas colecções e a anotação de conteúdos. Além disso, este projecto possibilita ainda a criação de ligações entre descrições de conteúdos usando ontologias e descritores visuais em formato MPEG-7 (MultiMedia, 2002).

Os resultados de investigação produzidos por este projecto ajudam os utilizadores a interagirem com os seus conteúdos multimédia através de tecnologias melhoradas de pesquisa e indexação, métodos automatizados de catalogação e, adaptação de conteúdos consoante os dispositivos e os cenários que o utilizador pretende. Garantem ainda a extensão e o enriquecimento de ontologias, de modo, a incluir, para além, das características semânticas, as características visuais que permitem descrever as propriedades sintácticas dos elementos multimédia.

Por sua vez, a plataforma CREAM - *CREating Relational Annotation-based Metadata* possibilita a criação de metadados relacionais de anotações para a Web semântica, assentando em métodos de anotação entre documentos anotados e, em anotações semânticas de recursos visuais baseada em ontologias. Esta plataforma vai também de encontro a um dos desafios mais interessantes nos dias que correm que é a captura de conhecimento sobre recursos multimédia digitais. Trata-se de pegar nos elementos sintácticos destes recursos e conseguir criar conhecimento estrutural interligado que permita representar informação relevante sobre esses elementos. Assim, é oferecido um ambiente para a construção de anotações na forma de metadados relacionais. Isto é, metadados que compreendem instâncias de anotações e relações entre instâncias de anotações que podem ser baseadas num determinado domínio ontológico e não numa estrutura ontológica fixa.

O M-OntoMat-Annotizer<sup>11</sup> é uma ferramenta cujo contributo se destina à aquisição de conhecimento para a anotação automática de conteúdos multimédia. Permite ao utilizador extrair descritores visuais de imagens e vídeos no formato MPEG-7 e, de seguida, armazená-los como modelos visuais de classes ontológicas. Estes modelos são armazenados como instâncias RDF usando uma versão RDF da norma para descritores MPEG-7 visuais.

A concretização do M-OntoMat-Annotizer (Bloehdorn et al., 2005) compreende uma componente de extracção visual de descritores (VDE) que apoia a transformação dos recursos multimédia XML extraídos, em instâncias de descritores visuais definidos no descritor visual ontológico (VDO). Para isso, é usada uma especificação de uma transformação XSL que cria a instância correspondente do descritor para cada uma das extracções e, que é depois, colocada na base de conhecimento. A VDE vincula ainda, de forma automática, a instância do descritor visual recém-criada, com a instância de protótipo de um conceito de domínio. Tanto as instâncias de protótipo de conceito, como os descritores extraídos

---

<sup>11</sup><http://mklab.itl.gr/m-onto2>

correspondentes, podem ser guardadas num arquivo RDF.

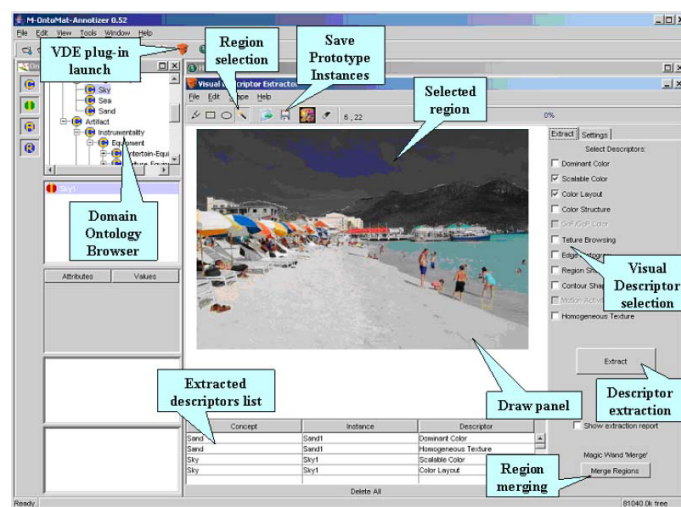


Figura 2.5: Anotação multimédia no M-OntoMat-Annotizer

Tal como a Figure 2.5 ilustra, a interface do M-Ontomat-Anotizer permite extrair recursos visuais, não só de imagens como um todo, mas também, no que respeita a objectos específicos incluídos numa imagem, o M-OntoMat-Annotizer permite ao utilizador desenhar uma região de interesse na imagem (e mesmo, numa imagem de uma trama de vídeo) e aplicar a extracção de descritores multimédia, apenas para essa região seleccionada, como pode ser observado na Figura 2.6. Como alternativa, para minimizar o esforço de anotação, o M-OntoMat-Annotizer também suporta a segmentação automática de imagens. Sempre que uma nova imagem é carregada, é automaticamente segmentada e dividida em regiões, deixando ao utilizador apenas a função de selecção das regiões desejadas. Para dar resposta à segmentação cruzada de regiões e ter mais de uma região à qual corresponde um único conceito semântico, a ferramenta permite ao utilizador juntar e agregar duas ou mais regiões, antes de este prosseguir com a extracção os descritores visuais. Note-se ainda que a finalidade do M-OntoMat-Annotizer é anotar imagens e imagens em vídeos para auxiliar o processo de análise multimédia *a posteriori*, e não produzir anotações para o utilizador final.

As linguagens suportadas para a ontologia de domínio incluem RDFS, DAML e OIL<sup>12</sup>. As anotações são produzidas ao nível da região, e a selecção de regiões pode ser realizada através desenho livre, de rectângulos, de elipses ou usando a tão conhecida varinha mágica. As anotação de vídeo não têm em conta a dimensão temporal, ou seja, a anotação é realizada trama a trama, em cada uma das tramas, seguindo o mesmo procedimento que para a anotação clássica de imagens. Para permitir localizar cada uma das representações, o M-OntoMat-Annotizer usa máscaras que ligam as respectivas instâncias, através de uma convenção de nomes.

<sup>12</sup><http://www.daml.org/2001/03/daml+oil-index.html>

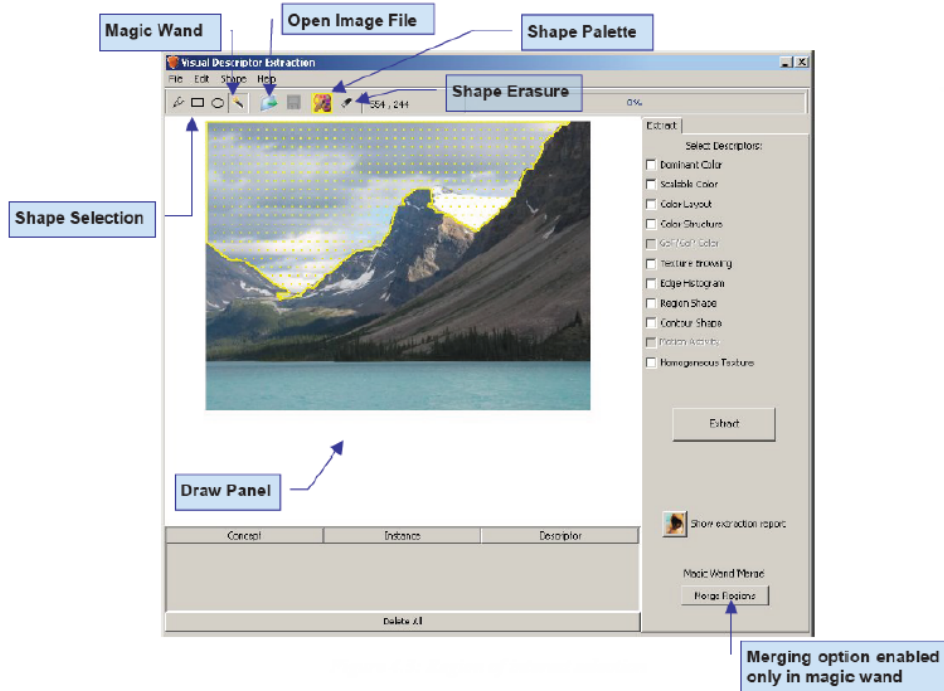


Figura 2.6: Extração de descritores visuais no M-OntoMat-Annotizer

## 2.2.4 Comparação de ferramentas

Nesta secção é apresentada a comparação entre as três ferramentas atrás mencionadas, com base na metodologia utilizada pelo *W3C Multimedia Semantics Incubator Group*<sup>13</sup>, no que respeita à avaliação de ferramentas de anotação semântica de imagens. Esta metodologia tem em conta, os seguintes critérios de comparação, usados na comparação de características-chave de ferramentas de anotação semântica de imagens:

- **Tipo de conteúdo:** as anotações podem incidir em diferentes tipos de conteúdo, quer em imagens quer em textos de páginas HTML, vídeos, e mesmo áudio;
- **Tipo de metadados:** cada anotação que se produz pode ser orientada para diferentes tipos de utilização, devendo estar categorizada através de metadados. Os metadados podem ser descritivos (para descrição e identificação de informação), estruturais (usado para descrever aspectos de navegação e apresentação) ou administrativos (para gestão e processamento da informação). No caso da anotação semântica de imagens pode ser muito útil, englobar no tipo de metadados descritivos, a vertente de descrição e identificação de características próprias de uma imagem (histograma de cores, texturas e formas geométricas), digamos que um conjunto de metadados *visuais* que são uma espécie de ADN da imagem;

<sup>13</sup><http://www.w3.org/2005/Incubator/mmsem/XGR-image-annotation>



- **Formato de metadados:** uma anotação pode ser expressa em diferentes formatos. O formato é importante pois deve garantir a interoperabilidade com outras aplicações. Por um lado, o MPEG-7 é frequentemente usado como o formato de metadados para o intercâmbio automático de resultados de análises de anotações semânticas. Por outro lado, o uso de RDF e OWL é considerado mais apropriado no mundo da Web Semântica;
- **Granularidade:** uma anotação é baseada numa imagem como um todo, ou em segmentos dessa imagem. É crucial conhecer a estrutura de uma imagem. Por exemplo, é útil obter anotações para diferentes regiões de uma imagem que permitam descrever a presença de informação na imagem (como por exemplo, que uma dada imagem contém uma parte textual ou que a sua estrutura, como imagem, resulta da composição de determinadas partes de imagens) ou ainda, definir e descrever objectos diferentes contidos na imagem;
- **Nível de anotação:** nas anotações de uma imagem podem ser usados vocabulários ou apenas texto livre. Quando são usadas ontologias, o nível de anotação é considerado controlado, já que a semântica da anotação é fornecida de uma forma formal e estruturada, ao passo que, se não for, o nível de anotação é considerado livre;
- **Persistência:** uma ferramenta de anotação deve ter em conta anotações realizadas anteriormente e, possuir a capacidade de reflectir isso mesmo, de um modo estruturado, apresentando não só essas anotações, mas também novas que venham a ser produzidas;
- **Controlo de acesso:** o tipo de acesso aos metadados distingue entre os utilizadores que apenas têm acesso para visualizar anotações e os outros que têm um acesso também de modificação de anotações;
- **Tipo de utilização:** existe uma plataforma Web colaborativa de anotação de imagens que disponibiliza a partilha da sua base de dados ou, apenas uma ferramenta de anotação de conteúdo multimédia para uso individual;
- **Requisitos de instalação:** a instalação de uma aplicação local ou o uso de um *browser* para aceder a um serviço Web de anotação de imagens e aí se realizarem operações de anotação sobre uma base de dados de imagens;
- **Licença:** o código é aberto ou não, e outras condições de licenciamento.

A Tabela 2.1 pretende resumir e comparar as características próprias das três ferramentas escolhidas para análise, ao longo da secção anterior, bem como, as tecnologias usadas e aplicadas por cada uma delas.

<b>Crêterios</b>	<b>Photostuff</b>	<b>AKTiveMedia</b>	<b>M-OntoMat-Annotizer</b>
tipo de conteúdos	imagem, RDF, HTML	imagem, texto, HTML	imagem, vídeo
tipo de metadados	descritivo, estrutural, administrativo	descritivo	descritivo, visual
formato de metadados	OWL	RDF	RDF
granularidade	imagem, região	imagem, região	imagem, região
nível de anotação	RDFS, OWL	RDFS, OWL, DAML	RDFS, DAML, OIL
persistência	sim	sim	sim
controle de acesso	inexistente	inexistente	inexistente
tipo de utilização	individual	individual	individual
requisitos de instalação	aplicação local	aplicação local	aplicação local
licenças	MPL <sup>a</sup>	GPL <sup>b</sup>	LGPL <sup>c</sup>

Tabela 2.1: Comparação entre ferramentas estudadas

<sup>a</sup><http://www.mozilla.org/MPL><sup>b</sup><http://www.gnu.org/copyleft/gpl.html><sup>c</sup><http://www.gnu.org/copyleft/lgpl.html>

## 2.3 Sumário

Este capítulo teve o seu foco na contextualização da ferramenta de anotação e classificação de imagens proposta nesta dissertação relativamente aos LFD e aos LFDE. Foram também, apresentados os aspectos relacionados quer com a reprodução, quer com a produção de LFDE e a integração dessa ferramenta no seio desta bancada de produção. Neste capítulo foram ainda descritas três ferramentas de anotação semântica, em particular, de anotação semântica de imagens. Foi possível identificar as características-chave dessas ferramentas, de modo a fornecer orientações para a criação de uma ferramenta de anotação semântica de imagens e sistemas de suporte, que melhor se adequasse ao contexto e aos requisitos exigidos pelo projecto *RiCoBA*.

A estratégia é agora, delineada de forma mais concreta, tendo como base a anotação de imagens de livros e a sua indexação num repositório multimédia. Esta estratégia passa pela construção de uma ferramenta que permita anotar imagens de diferentes formatos e, que possibilite, da mesma forma, anotar instâncias de imagens previamente anotadas, tal como todas as ferramentas analisadas neste capítulo, o fazem. À semelhança do Photos-tuff pode usar-se uma ontologia OWL fixa, de referência, no entanto, que seja estendida com novos conceitos, como a descrição de regiões em SVG e as descrições das partes visuais das regiões ou da própria imagem. Há ainda que considerar a interacção com um repositório multimédia, tanto na indexação, como na pesquisa de imagens e instâncias de anotações. Da mesma maneira, que o Photostuff, esta ferramenta deve ser capaz de produzir metadados descritivos e administrativos. No entanto, e à semelhança do M-OntoMat-Annotizer, deve englobar também, metadados visuais, tais como, os relativos ao histograma de cores, à cor dominante, e ainda, aos relacionados com a textura de uma determinada região ou da própria imagem.

É igualmente, importante que o processo de anotação se faça da forma mais livre possível, indo muito para além do AKTiveMedia, no sentido de permitir desenhar não só rectângulos e círculos, mas também, polígonos irregulares como acontece com o Photos-tuff. Possibilitar o desenho de formas elípticas e, de ter à disposição o uso da varinha mágica para definir contornos de regiões de interesse. Resumindo, ter formas semiautomáticas de anotação e mesmo, mecanismos automáticos, como é proporcionado pelo M-OntoMat-Annotizer, essencialmente, no que diz respeito à extracção de características visuais das imagens.



## Capítulo 3

# Desenho e Concretização

Neste capítulo, são apresentadas considerações sobre a arquitectura de classificação e gestão de conteúdos multimédia proposta, à luz da arquitectura de produção de LFDE, explanada no Capítulo 2. Com base nessas considerações, é proposta e descrita uma arquitectura genérica de processamento de conteúdos multimédia (PCM), capaz de lidar com a classificação e com a gestão, de diversos tipos de elementos multimédia diferenciados. Esta arquitectura é independente de contexto e de domínio de classificação; compreende, também, aspectos relacionados com a composição de elementos multimédia, nomeadamente, aspectos que tornam possível a sincronização ou a criação de uma qualquer relação entre elementos multimédia, quaisquer que sejam os seus tipos. Para além disto, inclui, ainda, um repositório e mecanismos de gestão de conteúdos multimédia. Ao longo deste capítulo, são ainda abordados os diferentes componentes da arquitectura e a relação entre estes, assim como os seus respectivos papéis nos processos de classificação e gestão de conteúdos multimédia.

Contudo, sendo o foco deste trabalho a classificação de conteúdos com suporte em imagens, propõe-se e apresenta-se, uma especialização da arquitectura genérica de PCM, para lidar exclusivamente, com elementos multimédia do tipo *imagem*, ao qual se chamou - unidade de processamento de imagens (UPI) .

### 3.1 Arquitectura genérica de PCM

Ao longo do processo de produção de LFD é importante garantir suporte ao enriquecimento de conteúdos originais. A necessidade de criar cenários diversificados de enriquecimento de LFD exige que a arquitectura de produção de LFDE contemple um repositório multimédia. Um repositório que armazene diversos tipos de elementos multimédia semanticamente indexados, de modo, a poderem ser pesquisados e extraídos sempre que necessário, com base em diversos critérios, tanto de base semântica, como da sua própria descrição sintáctica.

Torna-se, assim, necessário desencadear procedimentos de classificação de conteúdos.

Por um lado, criar mecanismos de extracção de características desses conteúdos, baseados em tarefas de anotação e identificação de regiões, sobretudo manuais, mas, sempre que possível, automatizadas; por outro lado, é necessário estabelecer representações coerentes, das características extraídas.

Apoiado nestes fundamentos, e, para fazer face aos aspectos descritos no Capítulo 2, na secção 2.1.3 *Processamento de Conteúdos* referentes à produção de LFDE, é proposta a arquitectura genérica de PCM, apresentada na Figura 3.1. Também nesta secção, segue-se a análise em mais detalhe de cada uma das componentes que constituem esta arquitectura.

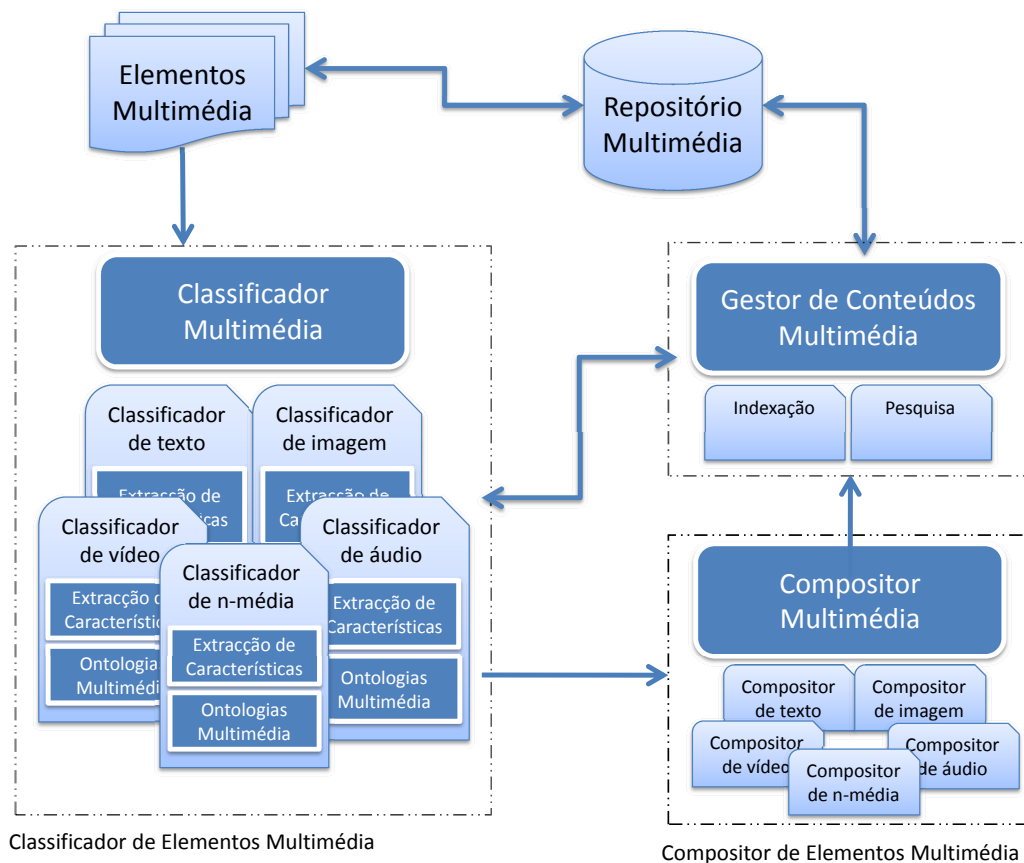


Figura 3.1: Arquitectura genérica de PCM

### 3.1.1 Classificadores multimédia

O componente de classificação de elementos multimédia agrega uma variedade de classificadores, cada um deles responsável pela classificação de um tipo específico de elemento multimédia, tal como texto, imagem, vídeo, etc. Perante um determinado tipo de elemento multimédia, cada classificador deve assumir a realização de duas tarefas: a extracção de características desse elemento e a representação dessas características sob a forma de ontologias multimédia. Na arquitectura de PCM apresentada, estas tarefas são

representadas através de um componente de extracção de características e de um outro de ontologias multimédia.

A primeira tarefa é levada a cabo pelo componente de extracção de características, responsável pela recolha, em diferentes níveis, de informação sobre um dado elemento multimédia, permitindo tanto a extracção de características estruturais ou sintácticas, como outras baseadas no seu conteúdo (semânticas). Esta tarefa é orientada em torno da categorização de texto, da compreensão de porções de uma imagem, da análise de um item auditivo, ou do estabelecimento de relações entre elementos constituintes de um dado elemento multimédia. A interpretação e anotação de elementos multimédia deve ser apoiada por instrumentos manuais, semiautomáticos ou, mesmo, automatizados, recorrendo para isso a algoritmos específicos para cada um dos tipos de elementos multimédia em análise. É, por este motivo, proposta uma abordagem centrada em mecanismos da *Web Semântica* para ir de encontro aos desafios de análise semântica e de anotação de elementos multimédia. Por exemplo, no caso da anotação de imagens podem ser empregues modelos de extracção de características padrão para a detecção de contornos e regiões na imagem, ou para a análise da textura da mesma.

Posteriormente, devem ser tomadas decisões relativamente à forma de representação e aos métodos de descrição das características extraídas, nomeadamente para descrever limites ou fronteiras, regiões e corpos sólidos identificados nos elementos multimédia.

Para que isto seja possível, a primeira tarefa, que tem como base o componente de extracção de características, terá de coexistir com uma segunda tarefa, que assente num componente de ontologias multimédia e, que forneça os meios adequados de representação para as anotações produzidas anteriormente. Assim, o componente de ontologias multimédia, através de métodos de criação ou de reutilização de ontologias multimédia, deve ser capaz de produzir representações de anotações que compreendam tanto descrições sintácticas, como descrições semânticas de um dado média, qualquer que seja o seu tipo. Espera-se ainda que os métodos seguidos estejam em conformidade com as mais actuais normas de representação de anotações de recursos multimédia e que, desta forma, possibilitem a criação de documentos semanticamente anotados pelo próprio autor. Devem, assim, ser criadas *etiquetas* semânticas derivadas de ontologias que descrevam, de forma clara, as características dos conteúdos dos média anotados. Para além disso, devem ser utilizadas adequadamente, novas ontologias para domínios específicos de um dado média, ao mesmo tempo que devem ser tidas em consideração ontologias previamente construídas - e importá-las, ou alargar o espectro das mesmas, isto é, estendê-las.

Uma vez mais, a iniciativa da *Web Semântica* é extremamente importante neste contexto já que fornece também modelos de representação de conhecimento para descrever diferentes tipos de recursos, nomeadamente, através de descritores de recursos RDF e de ontologias OWL, que podem ser aplicados, neste caso específico, aos recursos multimédia. Por exemplo, no que toca à classificação de imagens, a parte visual da norma

MPEG-7 ajuda a descrever e fazer relações entre as características dos média e os seus metadados. Consequentemente, o foco deve estar na definição de esquemas normalizados, de descrição de representações de características sintácticas e semânticas, dos próprios recursos multimédia. Esta normalização deve garantir interoperabilidade, facilitar a reutilização e assegurar a disponibilidade da informação entre as diversas ontologias multimédia.

O classificador de elementos multimédia, aqui descrito, também se encontra relacionado com o componente de composição de média e com o de gestão de conteúdos multimédia. Isto porque, por um lado, uma anotação, rica e formal que não é mais do que o resultado do classificador, servirá de entrada no caso de se pretender efectuar composições entre recursos já anotados; por outro lado, porque é importante que todas as anotações produzidas sejam colocadas no repositório multimédia, para que deste modo, possam servir a bancada de produção de LFDE.

### **3.1.2 Compositores multimédia**

O compositor de elementos multimédia compõe elementos multimédia, tendo em conta futuros cenários de enriquecimento de LFD, que tornem necessário conjugar entre si diversos elementos independentes. Pretende-se que seja capaz de combinar elementos multimédia simples, assim como elementos multimédia que já tenham sido alvo de uma qualquer forma de composição. Deste modo, o componente pode compor, combinar e/ou sobrepor elementos multimédia simples (p.ex., compor imagens com texto para formar um livro com imagens), assim como outros média, já compostos (p.ex., adicionar uma narração a um livro com imagens). As regras a usar no processo de composição devem ser definidas tendo em conta atributos simples baseados em metadados ou por meio de algoritmos complexos (p.ex., recorrendo a descrições semânticas complexas). Um caso prático que este componente deve ser capaz de retratar, prende-se com a capacidade de interligar imagens com texto e áudio previamente compostos, nomeadamente em livros falados digitais como descrito por Duarte et al. (2003).

De todo o processo de composição, deve resultar um conjunto sensível de informação composta que deve ser armazenada num repositório multimédia e estar disponível para ser usada ao longo do processo de criação de LDDE.

### **3.1.3 Gestor de conteúdos multimédia**

O último componente que caracteriza a arquitectura de PCM é o de gestão de conteúdos multimédia. Pretende-se, por um lado, que seja um componente de suporte às tarefas de indexação semântica e sintáctica de elementos multimédia, que potencie ou possibilite vários modos de indexação. Por outro lado, é desejável que providencie meios adequados e, acima de tudo, flexíveis e eficientes de pesquisa e extracção de informação, nomea-



damente baseada em conteúdos semanticamente indexados. Torna-se, então, condição necessária o armazenamento do resultado de toda e qualquer acção dos componentes atrás mencionados, num repositório multimédia. Este repositório deve ter capacidades de indexação que permitam indexar os próprios elementos multimédia, antes de serem alvo de processamento, por exemplo, uma imagem ou um vídeo, assim como a indexação das suas descrições estruturadas. Todavia, no que respeita à extracção de informação, os desafios são ainda mais ambiciosos e de uma complexidade bem maior, principalmente no que respeita à extracção de informação baseada em conteúdos. Por um lado, deve ser possível efectuar pesquisas simples baseadas no conteúdo anotado de um dado elemento multimédia. Por outro lado, deve ser possível executar pesquisas mais complexas, que passem pelo estabelecimento de relações semânticas entre conteúdos distintos de um leque de elementos multimédia, também estes distintos.

De modo a apoiar a construção de conteúdos para cenários de enriquecimento de LFD, o repositório multimédia deve estar disponível, tanto em tempo de produção, como em tempo de reprodução. Deve ainda, fornecer suporte a pesquisas e extracção de elementos multimédia, de modo a integrar-se em ferramentas de produção de conteúdo manuais e semiautomáticas. Sendo certo, que para que tal aconteça, os restantes componentes da arquitectura de PCM tem de ser capazes de assegurar as seguintes actividades:

- Identificação, classificação e organização contínua de elementos multimédia;
- Conversão e estruturação de dados num formato normalizado;
- Criação de relações entre os elementos multimédia, baseadas em diversos critérios, por exemplo, semânticos ou de composição.

## 3.2 Casos de Uso

De forma a explicar as soluções adoptadas para fazer face aos problemas levantados neste trabalho, serão descritos, de seguida, os casos de uso mais relevantes do protótipo proposto. Pretende-se assim, dar a conhecer os principais cenários de casos de uso do sistema que aqui se propõe. Na perspectiva do utilizador fica assim, mais claro o contexto em que o protótipo poderá ser utilizado, bem como os requisitos do mesmo para que possa ser funcional. Para se proceder à anotação e classificação semântica de imagens, é necessário, em primeiro lugar anotar regiões em imagens e, de seguida, classificar essas anotações descrevendo as regiões realizadas. A partir daqui, passa a ser possível, compor texto com anotações de imagens e fazer pesquisas sobre os conteúdos indexados. A Figura 3.2 apresenta os casos de uso de maior relevo que se prendem genericamente com a anotação e com a classificação de imagens; com a composição de anotações de imagens com texto e, com a pesquisa de imagens e de imagens anotadas. De seguida, são descritos

detalhadamente cada um destes casos de uso.

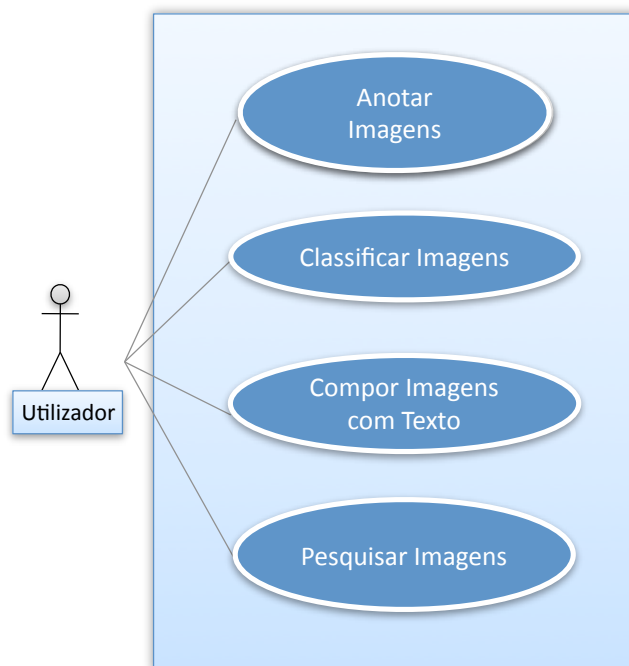


Figura 3.2: Caso de uso geral do protótipo

#### ***a) Anotar Imagens***

Este caso de uso compreende as funcionalidade presentes na Figura 3.3 e, que possibilitam ao utilizador iniciar o processo de criação de anotações de imagens.

##### *Abrir uma imagem para anotação*

Depois de se lançar a ferramenta, o utilizador tem a possibilidade de indicar anotar uma imagem que está guardada no seu computador ou, se preferir, escolher uma qualquer imagem presente no repositório multimédia e, até mesmo indicar uma que observe na *Internet*. Depois da selecção da imagem pretendida estar feita, esta é carregada para a área de desenho. Mais adiante nesta secção, é possível observar-se em mais pormenor, o processo de gestão de imagens no repositório multimédia. Existe um interface que permite ao utilizador indexar imagens, de modo, a que estas estejam disponíveis para serem usadas no processo de anotação.

##### *Abrir uma anotação*

Este caso de uso prende-se com a abertura de uma anotação, possibilitando ao utilizador, assim como, o anterior caso de uso iniciar o processo de criação de anotações. Deste modo, o utilizador tem agora de indicar, não uma simples imagem, mas uma

anotação de uma imagem previamente anotada. O utilizador pode indicar uma qualquer anotação presente no repositório multimédia que dá suporte a todo o processo de anotação e classificação de imagens, ou, se preferir, uma qualquer anotação que esteja guardada numa pasta no seu computador local. Depois de seleccionada a anotação pretendida, esta é carregada para a área de trabalho.

#### *Realizar anotação*

Este caso de uso, descreve como se realiza uma anotação, após a abertura de uma imagem ou de uma anotação, isto é, depois de se dar início ao processo de anotação. O sistema possibilita a escolha de um conjunto de ferramentas para apoio ao desenho de regiões. Assim, o utilizador começa por delimitar regiões de interesse na imagem e, em seguida tem à sua disposição, um conjunto de ferramentas específicas para a detecção e extracção de características sintácticas, nas regiões demarcadas, que ele próprio, acabou de compilar. São exemplos destas ferramentas a detecção de texturas e cores de regiões da imagem ou da própria imagem como um todo. Pode ainda, descrever do ponto de vista semântico as características que representam as regiões delimitadas.

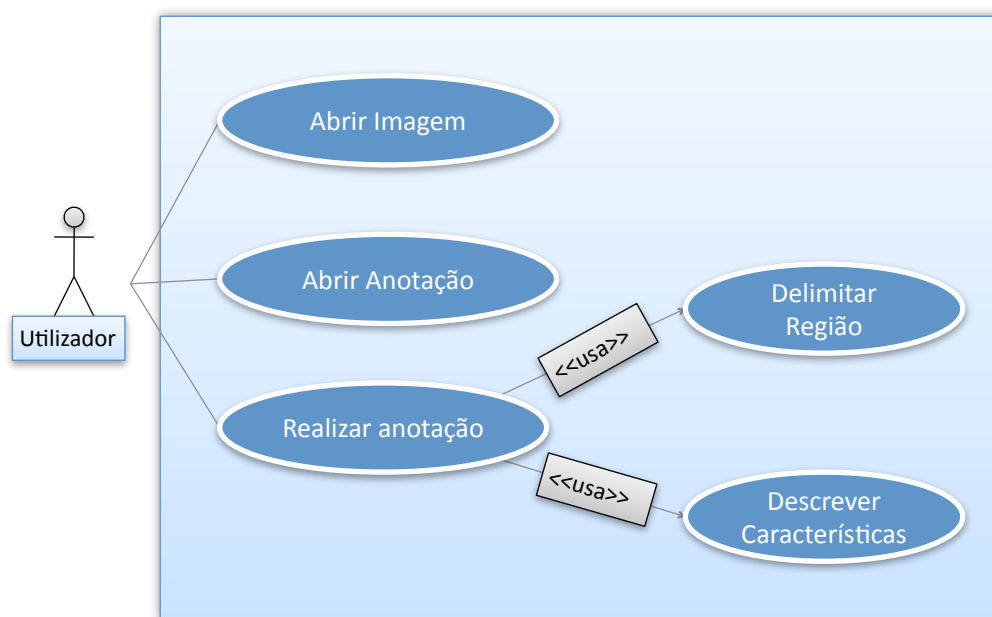


Figura 3.3: Caso de uso anotação de imagens

#### ***b) Classificar Imagens***

Este caso de uso possibilita ao utilizador iniciar o processo de classificação de anotações que produziu na imagem em estudo. Depois de realizada a anotação, o utilizador tem ao seu dispor funcionalidades de representação semântica e de etiquetização de regiões anotadas. De forma, a se representar a imagem como um todo, existem etiquetas como o

título, ou palavras-chave que possibilitam descrever a imagem. Para que se possa representar cada uma das regiões anotadas, existem etiquetas como o nome e a descrição da anotação, ou o tipo de polígono que representa a região anotada. Por exemplo, um círculo envolvendo o sol numa dada imagem.

#### *c) Compor Imagens com Texto*

Este caso de uso descreve a funcionalidade de composição de anotações específicas de uma determinada imagem ou da própria imagem com texto proveniente de um livro digital falado com. O utilizador carrega o texto de um ficheiro que representa um LFD e, escolhe um bloco de texto, um parágrafo, um capítulo, ou um conjunto de palavras com as quais pretende relacionar com uma dada anotação ou com uma determinada imagem.

#### *d) Pesquisa de Imagens*

Este caso de uso descreve as operações que são necessárias desencadear para que se possa pesquisar imagens e anotações de imagens. Deste caso de uso faz parte tanto a pesquisa de imagens guardadas na base de dados, como as anotações das mesmas. Deve ser possível pesquisar uma imagem, abri-la e, tratar de anotá-la e classificá-la, quer tenha sido anotada no passado, quer seja a primeira vez que pretenda anotar algo nessa imagem.

### **3.3 Desenho do protótipo**

As escolhas sobre as tecnologias a usar na construção da ferramenta de suporte à classificação de imagens, passaram sobretudo por encontrar o melhor ambiente computacional capaz de pôr em prática os objectivos de anotação e classificação de imagens digitais. Optando-se pelo *MATLAB* por ser, uma linguagem de alto nível bastante intuitiva, com um ambiente computacional bastante flexível e eficiente, em que se destaca a versatilidade no processamento e visualização de imagem digital. Também, porque o ambiente de desenvolvimento que o *MATLAB* fornece, apresenta garantias de eficiência em processamento de dados vectoriais e matriciais e, em particular, em processamento de imagem digital. Para além disso, a vasta e completa colecção de bibliotecas que dispõe, alarga o espectro de escolhas de entre os diversos algoritmos de processamento de imagem que se pretende usar. Assim, tanto é possível levar a cabo tarefas simples, como outras de maior complexidade que exijam bastante poder computacional.

Foi necessário, ainda, encontrar soluções para representar a classificação das imagens anotadas. Por fim, achar formas eficientes e robustas, tanto de indexação como de pesquisa, dessa representação, no repositório multimédia.

O protótipo aqui apresentado, como se pode observar na Figura 3.4 refere-se à fer-

ramenta baptizada de SICA, cujo foco principal é a anotação e classificação semântica de imagens. No entanto, esta ferramenta trata também aspectos relacionados com a composição de imagens com texto, e ainda, com indexação e pesquisa de imagens. Pretende-se nesta secção dar a conhecer, de forma detalhada, os diversos componentes que constituem esta ferramenta, o seu propósito, e o modo, como estes componentes se interligam entre si. A interface com o utilizador (UI) apresentado na Figura 3.5 retrata a generalidade dos espaços da interface gráfico que o utilizador tem ao seu dispor. De seguida, apresentam-se detalhadamente, as componentes gerais que constituem o protótipo.

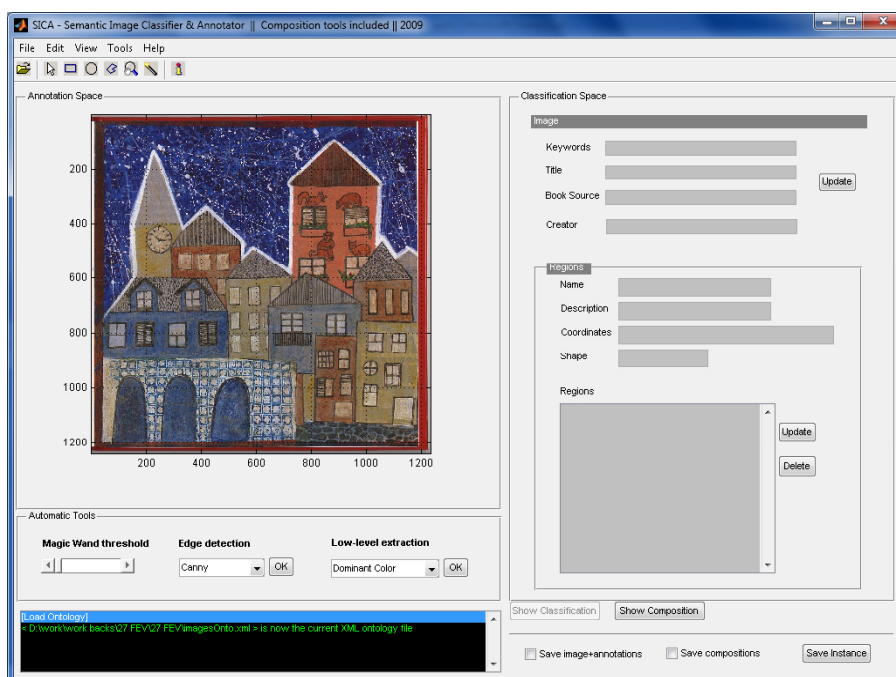


Figura 3.4: Vista da Ferramenta

### ***Menu e barra de ferramentas***

Na Figura 3.6 pode observar-se o *Menu* do protótipo, à semelhança da grande maioria dos protótipos de software que se constroem. É a partir do *Menu* que se desencadeiam os processos necessários à anotação e classificação semântica de imagens. Para além do *Menu*, o protótipo dispõe também, de uma *Barra de ferramentas* que se pretende ser essencialmente dedicada às tarefas de anotação.

Para além, dos componentes referido atrás, o protótipo apresenta componentes chave, que consistem em três grandes espaços de trabalho: o espaço de anotação, o espaço de classificação e, o de composição de imagem. Estes espaços, tal como o nome indica, são onde se desenvolvem as tarefas específicas de cada um dos componentes, de acordo com a arquitectura de PCM apresentado anteriormente, no início deste capítulo, aqui aplicado em relação a imagens e à sua composição com elementos textuais.

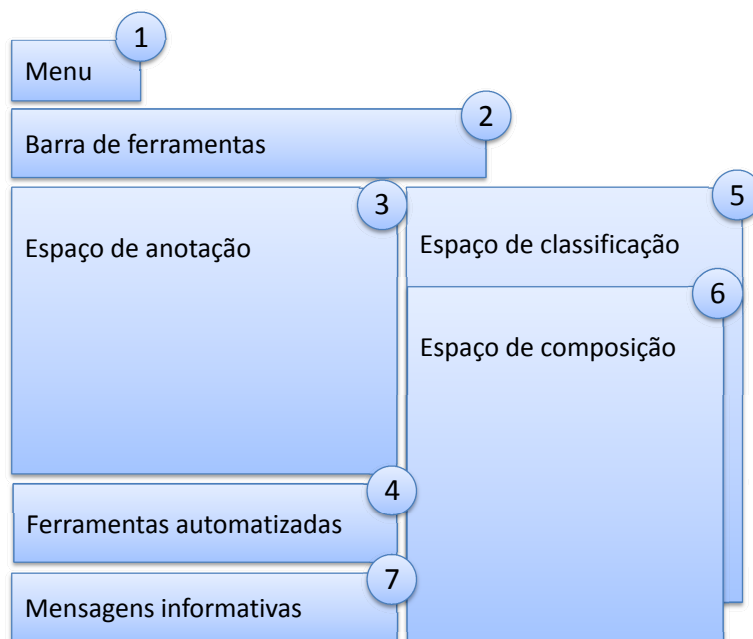


Figura 3.5: Interface com o utilizador



Figura 3.6: Menu e barra de ferramentas

### ***Espaço de anotação***

O espaço de anotação apresentado na Figura 3.7 é onde se desenvolvem as tarefas relacionadas com as imagens propriamente ditas, suportadas pelas ferramentas mencionadas atrás. Neste espaço, é possível, efectuar operações sobre imagens que se pretende anotar e, delinear regiões sobre essa imagem, para as quais, de seguida, se propõe atribuir uma determinada classificação semântica e/ou efectuar composições com textos de um livro. Um processo de anotação que envolve a utilização da barra de ferramentas pressupõe o desenho e a delineação manual de formas que envolvam as regiões de interesse de uma imagem. De forma, a complementar estes pressupostos está também disponível outros mecanismos, estes agora, automáticos que permitem identificarem e delinear regiões de interesse pretendidas.

### ***Ferramentas automatizadas***

A Figura 3.7 mostra ainda a área de ferramentas automatizadas que coloca ao dispor do utilizador ferramentas como a varinha mágica, e instrumentos para a segmentação e extracção de características visuais de uma imagem. Tendo uma ligação estreita com o espaço de anotação, o uso deste tipo de mecanismos automatizados, torna possível, sim-

plificar e agilizar as operações sobre imagens. Por exemplo, com a utilização da varinha mágica deixa de ser necessário a intervenção do utilizador na criação e delineamento por-menorizados e completos de formas de regiões de interesse numa dada parte da imagem. Do mesmo modo, que aplicando a ferramenta de segmentação de imagem, a uma imagem no seu todo, é possível, detectar e adquirir de forma automática, um conjunto de contornos e regiões da imagem. É ainda exequível extrair informação sobre as características visuais (ex., mapa de cores, cor dominante e texturas) de uma imagem, ou mesmo de partes ou regiões de uma imagem de forma automática.



Figura 3.7: Espaço de anotação e ferramentas automatizadas

### *Espaço de classificação*

O espaço de classificação, tal como é apresentado na Figura 3.8 é o espaço destinado à descrição de informação respeitante à imagem que se está a anotar. Assim, é possível efectuar o preenchimento de uma série de descritores que servem de apoio à descrição semântica de uma dada imagem em análise. Sugerir um título para a imagem ou associar uma imagem a uma fonte, ou a um nome de um livro são exemplos destes descritores de imagem. No que respeita à descrição das regiões, existe um gestor de regiões anotadas que permite lidar com descritores, tanto ao nível das características sintácticas, como das semânticas, próprias de cada região. Alguns exemplos destes descritores de regiões são a descrição e a forma de uma qualquer região de interesse.

Este espaço está inserido na mesma área de trabalho que o espaço de composição descrito de seguida. Assim, depois de se delinear regiões, é possível, alternar entre um espaço e o outro, ora para se classificarem regiões ora para se compor essas regiões ano-

tadas com texto. Esta sobreposição de espaços permite que o utilizador não perca o foco tanto das regiões anotadas na imagem que está a processar, quanto da classificação que fez até então, de cada região anotada.

The screenshot shows a web-based interface titled "Classification Space". It features a top navigation bar with a blue circle containing the number "5". Below the bar, there are two main sections. The first section, labeled "Image", contains four input fields: "Keywords", "Title", "Book Source", and "Creator", followed by an "Update" button. The second section, labeled "Regions", contains four input fields: "Name", "Description", "Coordinates", and "Shape", followed by a list of "Regions" with "Update" and "Delete" buttons. At the bottom of the interface, there are two buttons: "Show Classification" and "Show Composition".

Figura 3.8: Espaço de classificação

### *Espaço de composição*

A Figura 3.9 mostra o espaço onde é possível compor texto com imagens anotadas. Para tal, é necessário, por um lado, carregar no espaço de anotação, uma imagem simples ou uma imagem previamente anotada, e por outro lado, lançar no espaço de composição o texto de um livro, proveniente de um ficheiro estruturado que represente a estrutura de um LFD. O utilizador tem também ao seu dispor a possibilidade de alternar entre o espaço de classificação e o de composição, sem perda de contextualização. Esta interligação de espaços permite anotar regiões e, simultaneamente, compor as mesmas com blocos de texto.

Depois de explanado o desenho do protótipo e as suas componentes de IU, é agora tempo, de focar o modo como cada um dos componentes da arquitectura de PCM é con-



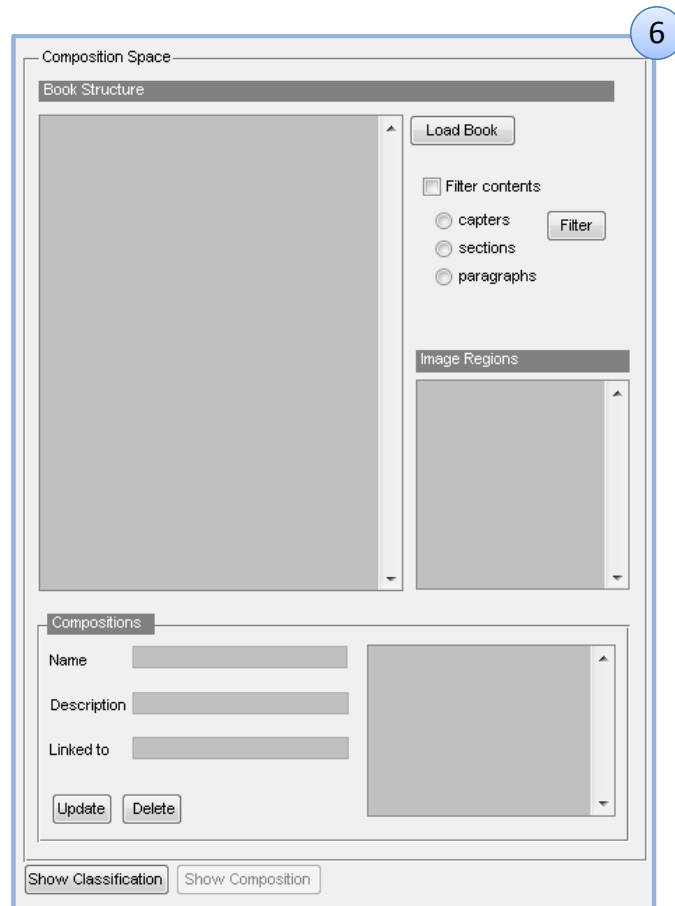


Figura 3.9: Espaço de composição

cretizado. Desta forma, mostrar também, como é possível ser efectuado um ciclo completo de processamento de conteúdos baseados em imagem, e a partir daí, demonstrar a sua aplicabilidade a qualquer tipo de conteúdos multimédia.

### 3.4 Unidade de processamento de imagem

Tendo como base a arquitectura de PCM anteriormente descrita e, sendo conhecida a necessidade de encontrar uma solução eficaz para a classificação de conteúdos com suporte em imagens, o foco deste trabalho assenta, pois, numa unidade de processamento de imagens, como pode ser observado na Figura 3.10. Esta UPI prevê a criação de uma ferramenta de classificação e indexação de imagens, capaz de pôr cobro, às lacunas, até agora, existentes no projecto *RiCoBa*. Assim, passará a ser possível efectuar anotações semânticas em imagens; combinar imagens ou regiões de imagens com blocos de texto, e vice-versa; e ainda, pesquisar de forma eficiente elementos multimédia não processados, ou compostos, no repositório multimédia que dá suporte à UPI. Deste modo, é garantida e validada a arquitectura de PCM, para o caso das imagens, na sequência do que foi feito,

em relação, à classificação e indexação de elementos áudio, no trabalho desenvolvido por Caseiro et al. (2002), na elaboração do alinhador do *INESC-ID*<sup>1</sup>, ainda no âmbito do projecto *IPSOM*.

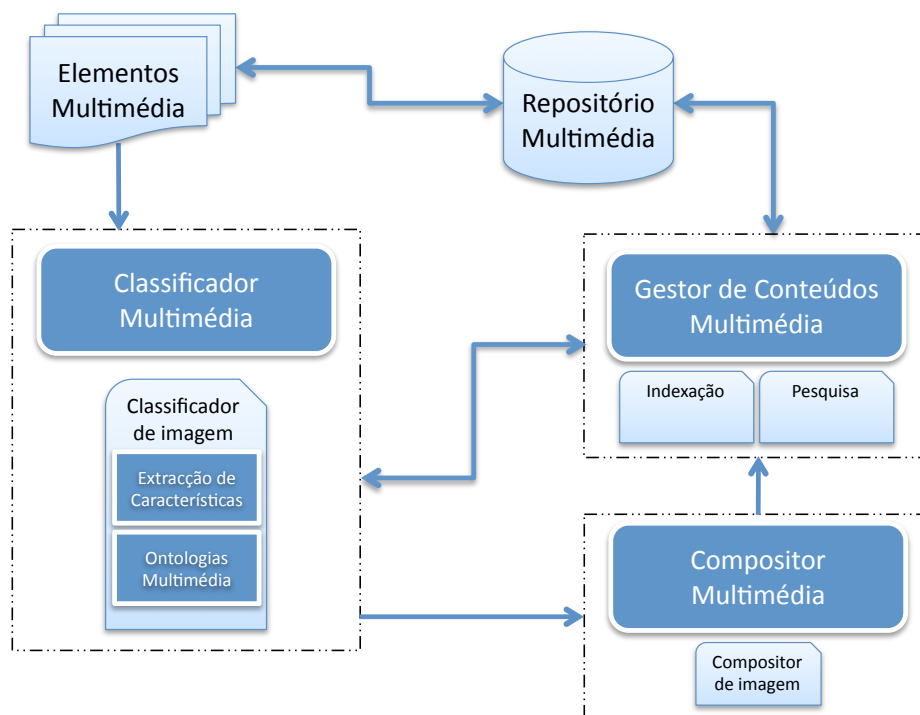


Figura 3.10: Arquitectura da UPI

### 3.4.1 Extracção de características

Esta subsecção pretende traçar de forma detalhada a concretização das funcionalidades associadas ao componente de extracção de características, tal como ilustrado na Figura 3.11, que é parte integrante do componente classificador de imagem, referente à arquitectura da UPI. Para dar início ao processo de extracção de características de conteúdos baseados em imagens, é necessário, em primeira instância, efectivar procedimentos de desenho e etiquetagem em imagens. Partindo, assim, de uma, ou de um conjunto de imagens com formas delineadas e, devidamente anotadas, pretende-se, de seguida, que sejam extraídas características próprias de partes ou regiões de interesse de cada imagem, ou mesmo, características da imagem como um todo. De facto, trata-se acima de tudo, de efectuar anotações sintácticas em imagens.

O espaço de anotação, tal como apresentado na Figura 3.12 serve para apoiar o utilizador nas tarefas de desenho e etiquetagem. Durante estas tarefas realizam-se operações

<sup>1</sup><https://www.l2f.inesc-id.pt>

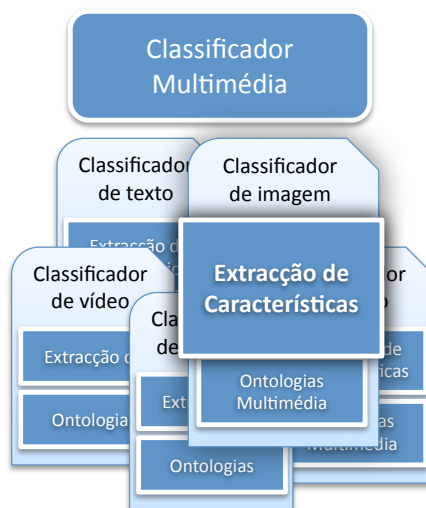


Figura 3.11: Componente de extração de características

que podem ser efectivadas através da utilização, quer da barra de ferramentas quer das ferramentas automatizadas, descritas anteriormente. A concretização deste conjunto de operações tem como requisitos, facilitar e simplificar a sua execução através do manuseamento das respectivas ferramentas. Por isso, umas operações são realizadas com o auxílio de ferramentas que precisam de intervenção manual com o utilizador e, outras são executadas automaticamente. Ainda assim, existe uma complementaridade entre estes tipos de operações, já que as operações manuais podem ser o ponto de partida para desencadear, se necessário, operações automáticas de extração de características.

#### ***Descrição de operações manuais sobre imagens***

O desenho manual e livre de regiões ou de áreas de desenho sobre a imagem em foco pode ser feita das seguintes formas:

- 1) usando formas rectangulares e elípticas;
- 2) usando polígonos livres, que corresponde ao desenho de formas irregulares e livres, mas fechadas, como pode ser observado na Figura 3.13. A concretização desta operação tem como base um ponto de partida, e de seguida, um conjunto de pontos que correspondem a diversos cliques ao longo de uma região de interesse. O polígono fica completo e fechado automaticamente devido ao cálculo de uma última linha, que une o último ponto, ao ponto inicial.

#### ***Descrição de operações automáticas sobre imagens***

O desenho automático de regiões ou de áreas de desenho sobre a imagem em foco, e a extração automática de características visuais de uma imagem, pode ser feito de três formas distintas:

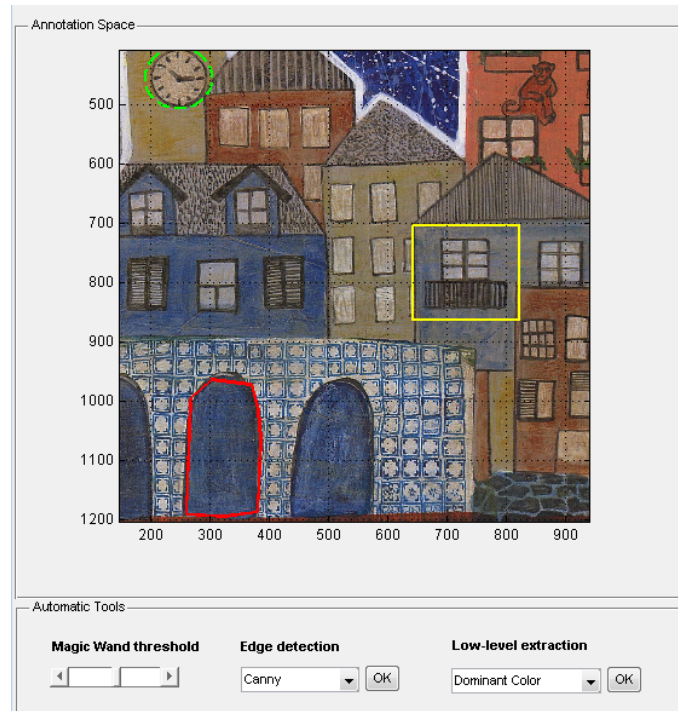


Figura 3.12: Anotação de imagens

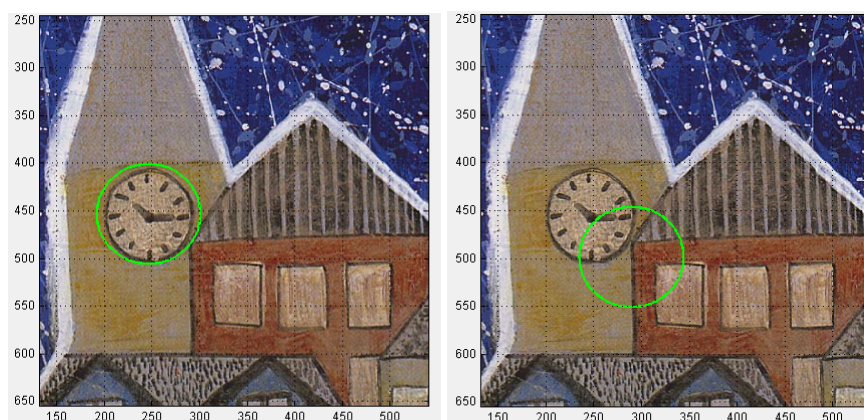
- 1) usando a varinha mágica para delinear automaticamente desenhos, ao longo de pontos vizinhos que possuem a mesma concordância de cor. Permitindo assim, criar formas que se traduzem em regiões, junto da fronteira de uma determinada região seleccionada. Para apoiar esta tarefa existe uma escala que deve ser configurada e, que ajuda a aumentar ou diminuir o espectro dos pontos vizinhos relacionados com as tonalidades de cor definidas à partida;
- 2) através das operações de segmentação de imagem que permitem a detecção de contornos de forma automática recorrendo a diversos algoritmos específicos para esse fim (p. ex., *canny* ou *sobel*). O objectivo desta operação é pois, identificar numa imagem, um conjunto de formas que delimitem objectos e regiões. Sendo assim possível, identificar e criar de uma só vez, diversas formas e regiões automaticamente numa imagem;
- 3) através da extracção de características visuais, utilizando técnicas de detecção de discontinuidades: detecção de linhas, contornos, ligação de contornos e detecção de fronteiras, para se extrair, de modo automático, características de uma imagem ou de regiões previamente identificadas e delineadas. As características que são possíveis de extrair têm que ver, com a cor dominante, a textura e o histograma de cores.

### ***Descrição de operações complementares***



Figura 3.13: Desenho livre

Tal como apresentado na Figura 3.14 existem outras operações que auxiliam o utilizador na execução de operações de anotação. Estas operações incluem a selecção e deslocação de regiões anotadas; o aumento e diminuição de *zoom*, assim como, a deslocação de uma região anotada depois de aumentada ou diminuída. Deste modo, melhoram o grau de precisão no desenho de regiões e formas numa imagem.



### *Automação de operações*

A Tabela 3.1 descreve o grau de automação das operações que acabaram de ser des-



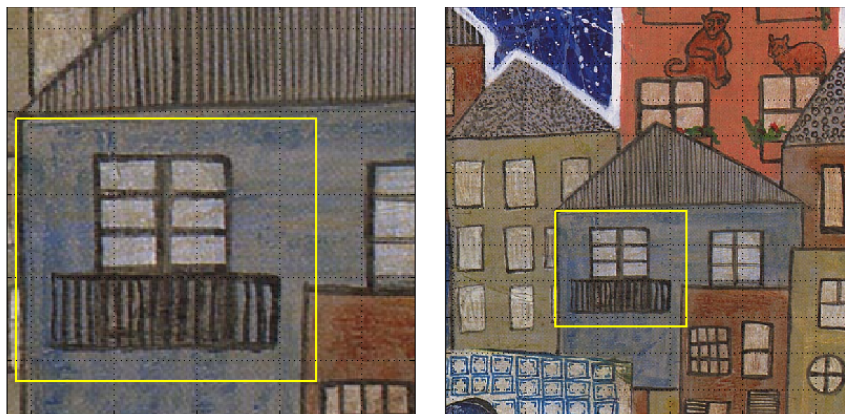


Figura 3.14: Operações de selecção e deslocamento e, de zoom

critas. A automação visa garantir melhores desempenhos de qualidade, no processo de extracção de características de imagens.

<b>Operações</b>	<b>Manuais</b>	<b>Automáticas</b>
Desenho de rectângulos, elipses e polígonos	X	
Seleccionar e mover	X	
Zoom e pan	X	
Varinha mágica		X
Segmentação de imagem		X
Extracção de características <i>visuais</i>		X

Tabela 3.1: Grau de automatização de operações sobre imagem

Depois de extraídas as características de uma imagem e, de se anotar devidamente essas características, torna-se então, necessário representá-las através de ontologias multimédia. Nesta subsecção foram apresentados alguns exemplos que exprimem o comportamento das ferramentas de extracção de características. No entanto, é no Capítulo 4 que pode ser observado, em detalhe, a articulação entre as diferentes ferramentas, operações e procedimentos aqui apresentadas.

### 3.4.2 Ontologias multimédia

Nesta subsecção é traçada de forma detalhada, a concretização das funcionalidades associadas ao componente de ontologias multimédia, que a par do componente de extracção de características, tornam efectiva a concretização do componente classificador de imagem, referente à arquitectura da UPI, ilustrada na Figura 3.15. O uso de uma ontologia multimédia genérica, não só satisfaz o propósito de representar anotações de imagens, como garante ainda espaço para a representação de anotações de outros tipos de elementos multimédia. Portanto, as representações oriundas de diversos classificadores de *média*, apesar



Classification Space

Image

Keywords: Lisboa, Sé, Alfama, Igrejas, Monumento

Title: Lisboa à noite

Book Source: Cidades portuguesas

Creator: Hugo Simões

Update

Regions

Name: Relógio

Description: relógio na fachada da igreja de

Coordinates: [189.3172 404.8945 114.2273 101.9887]

Shape: Ellipse

Regions

Relógio

Tunel

Varanda

Update

Delete

Show Classification

Show Composition

Figura 3.16: Classificação de imagens

- o autor e a fonte do livro a que pertence a imagem que se está a anotar;
- edição do nome e da descrição de cada região;
- detalhes das coordenadas que representam o traçado da região, ou seja, a localização da região na imagem;
- a forma e, a listagem de todas as regiões de interesse anotadas;
- remoção automática de regiões seleccionadas da lista de regiões ou da imagem, que se está a anotar;
- navegação na lista de regiões e visualização das regiões já anotadas e, das suas descrições semânticas;
- escolha de regiões a partir da listagem, sendo que a sua selecção, fica realçada na imagem, através de efeitos ou mudança de cor da forma anotada;
- movimentar uma anotação seleccionada para outro local da imagem, e actualizar as suas coordenadas automaticamente.



O facto, de se utilizar, um tipo de representação baseado em ontologias multimédia permite que a representação de uma dada anotação esteja numa forma normalizada e, por isso, facilite o intercâmbio de meta informação entre as diversas ferramentas que compõem o projecto. A flexibilidade desta representação permite também, que seja extensível a outros tipos de elementos multimédia que se pretenda anotar e representar, para além das imagens.

O processo de representação de imagens consiste na agregação de metadados associados à imagem e às suas partes. Em relação à imagem na sua globalidade, estão disponíveis diversos campos que correspondem a metadados descritivos e administrativos, que seguem normas, na representação de imagens anotadas. Partindo da norma *Dublin Core* temos uma representação da imagem como um todo. As regiões anotadas são representadas usando uma ontologia de imagens produzida em OWL+RDF.

De seguida, é apresentada a ontologia de imagens genérica que é usada na representação semântica e sintáctica, de cada uma das anotações de imagens produzidas pela ferramenta *SICA*, aqui proposta.

---

```
<?xml version="1.0"?>

<rdf:RDF
  xmlns:dc= "http://purl.org/dc/elements/1.1/"
  xmlns:h.0="http://www.w3.org/2003/12/exif/ns/"
  xmlns:h.1="http://www.mindswap.org/~glapizco/technical.owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:h.2="http://www.w3.org/2004/02/image-regions#"
  xmlns:svg="http://www.w3.org/2000/svg" >

  <rdf:Description rdf:type="http://www.mindswap.org/~glapizco/technical.owl#Image" t:←
    tags="" rdf:about="" >

    <dc:title>A name given to the resource</dc:title>
    <dc:creator>An entity primarily responsible for making the content of the resource</dc:←
      creator>
    <dc:subject>A topic of the content of the resource</dc:subject>
    <dc:description>An account of the content of the resource</dc:description>
    <dc:publisher>An entity responsible for making the resource available</dc:publisher>
    <dc:contributor>An entity responsible for making contributions to the content of the←
      resource</dc:contributor>
    <dc:date>A date of an event in the lifecycle of the resource</dc:date>
    <dc:type>The nature or genre of the content of the resource</dc:type>
    <dc:format>The physical or digital manifestation of the resource/html</dc:format>
    <dc:identifier>An unambiguous reference to the resource within a given context</dc:←
      identifier>
    <dc:source>A Reference to a resource from which the present resource is derived</dc:←
      source>
    <dc:language>A language of the intellectual content of the resource</dc:language>
    <dc:relation>A reference to a related resource</dc:relation>
    <dc:coverage>The extent or scope of the content of the resource</dc:coverage>
    <dc:rights>Information about rights held in and over the resource</dc:rights>

    <h.0:imageLength>Image length</h.0:imageLength>
    <h.0:imageWidth>Image width</h.0:imageWidth>

    <h.1:hasRegion rdf:resource=""/>
```

```

</rdf:Description>

<rdf:Description rdf:type="http://www.mindswap.org/~glapizco/technical.owl#ImagePart" ␣
  rdf:about="">

  <h.2:coords>Region coordinates</h.2:coords>
  <rdf:type rdf:resource=""/>
  <rdfs:label>Region Label</rdfs:label>
  <h.1:regionOf rdf:resource=""/>

  <h.1:svgOutline>
  <svg:svg preserveAspectRatio="none" width="4in" height="3in" version="1.1"
  xmlns:xlink="http://www.w3.org/1999/xlink"/>

  <desc></desc>

  <image xlink:href="" x="0" y="0" width="" height="" />

  <rect x="" y="" width="" height="" style="fill:none; stroke:yellow; stroke-width:2pt;␣
  ;" />

  <ellipse cx="" cy="" rx="" ry="" style="fill:none; stroke:yellow; stroke-width:2pt;"␣
  />

  <polyline points="" style="fill:none; stroke:blue; stroke-width:1pt;" />

  </svg:svg>
  </h.1:svgOutline>
  <rdf:type rdf:resource="http://www.mindswap.org/~glapizco/technical.owl#ImagePart"/>

</rdf:Description>

<rdf:Description rdf:about="">
<owl:imports rdf:resource="http://www.mindswap.org/~glapizco/technical.owl#"/>
<rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Ontology"/>
</rdf:Description>

</rdf:RDF>

```

---



---

Listing 3.1: Ontologia de imagens genérica

Para manipular as estruturas ontológicas criadas foram construídas diversas funções que tiram proveito, do facto, das anotações estarem guardadas em formato XML. Para além disto, como as anotações das diversas regiões de uma imagem estão representadas em RDF, possibilitam que se pegue em cada um dos ficheiros RDF produzidos e se arraste para dentro de um *browser*, a fim de se obter a visualização de cada uma das anotações. Isto é só possível porque o conteúdo deste RDF possui um pedaço de descrição baseado em SVG, fazendo com que o tratamento dado pelo *browser* seja semelhante a um ficheiro SVG. Para manipulação de estruturas *XML* usou-se a biblioteca *XMLTree* (Guillaume Flandin, 2003) realizada em *MATLAB* e *C*.

### 3.4.3 Composição

A Figura 3.17 mostra o componente de composição de imagem, da arquitectura de UPI cuja concretização é apresentada nesta subsecção. Tal como o próprio nome indica, a concretização de um compositor de imagem tem como objectivo principal permitir a composição de imagens com outros tipos de média. Neste caso, em particular, dado que

esta dissertação propõe a validação da arquitectura de PCM, tendo como base as imagens de um livro, a concretização do componente de composição, passa necessariamente, pelo estabelecimento de relações de composição entre imagens e texto de um livro.

Sendo o texto, o principal foco de composição do componente explanado nesta secção, é necessário, encontrar uma forma de criar interligações entre anotações de imagens e partes de texto de um qualquer livro. Assim, com base num conjunto de parágrafos, secções e capítulos de um dado texto de um livro, deve ser possível, criar e enumerar relações com diversas anotações de imagens ou imagens como um topo. Para tal, o uso dos dois componentes de classificação de imagens apresentados nas duas secções anteriores é a base para a criação destas relações, em torno das anotações realizadas.

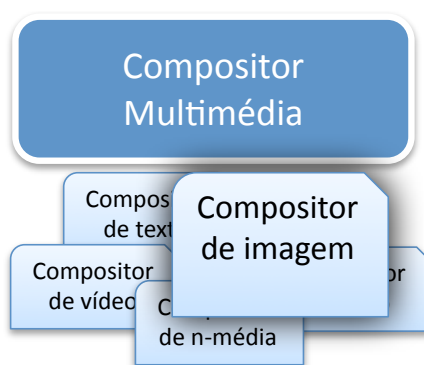


Figura 3.17: Componente de composição de imagem

O espaço de composição ilustrado na Figura 3.18 apresenta as características do componente de composição de imagem, assim como, as operações que dão suporte às tarefas de composição de imagens anotadas. O componente compositor de imagem é constituído, por um gestor de estruturas de textos e por um gestor de composições. Em primeiro lugar, os textos de um livro devem ser carregados para o espaço de composição. Depois de carregado todo o texto, é possível, visualizar toda a sua estrutura e filtrar o seu conteúdo por parágrafo, secção ou capítulo. Para além disso, existe ainda uma lista de anotações de imagem que devem ser mapeadas com o texto, bastando para isso associar estas, a blocos de texto apresentado no gestor de estruturas de texto. A criação de uma nova composição de anotações de imagens com blocos de texto torna-se, de facto efectiva, após a criação de um nome e de uma descrição, e da selecção do bloco de texto e da anotação de imagem, que se pretende mapear.

---

```
<?xml version="1.0" ?>
<Book>
<chapter id="1" rdf:resource="...">Capítulo I
  <section rdf:resource="...">
    <paragraph rdf:resource="...">
      <phrase rdf:resource="...">O João come a papa.</phrase>
      <phrase rdf:resource="...">O Francisco também.</phrase>
```

```

</paragraph>
</section>
</chapter>
<chapter id="2" rdf:resource="...">Capítulo II
  <section rdf:resource="...">
    <paragraph rdf:resource="...">
      <phrase rdf:resource="...">O Hugo come a papa.</phrase>
      <phrase rdf:resource="...">O Rui também.</phrase>
    </paragraph>
  </section>
</chapter>
<chapter id="3" rdf:resource="...">Capítulo III
  <section rdf:resource="...">
    <paragraph rdf:resource="...">
      <phrase rdf:resource="...">O Pedro come a papa.</phrase>
      <phrase rdf:resource="...">O Afonso também.</phrase>
    </paragraph>
  </section>
</chapter>
</Book>

```

Listing 3.2: Representação da estrutura XML de um LFD

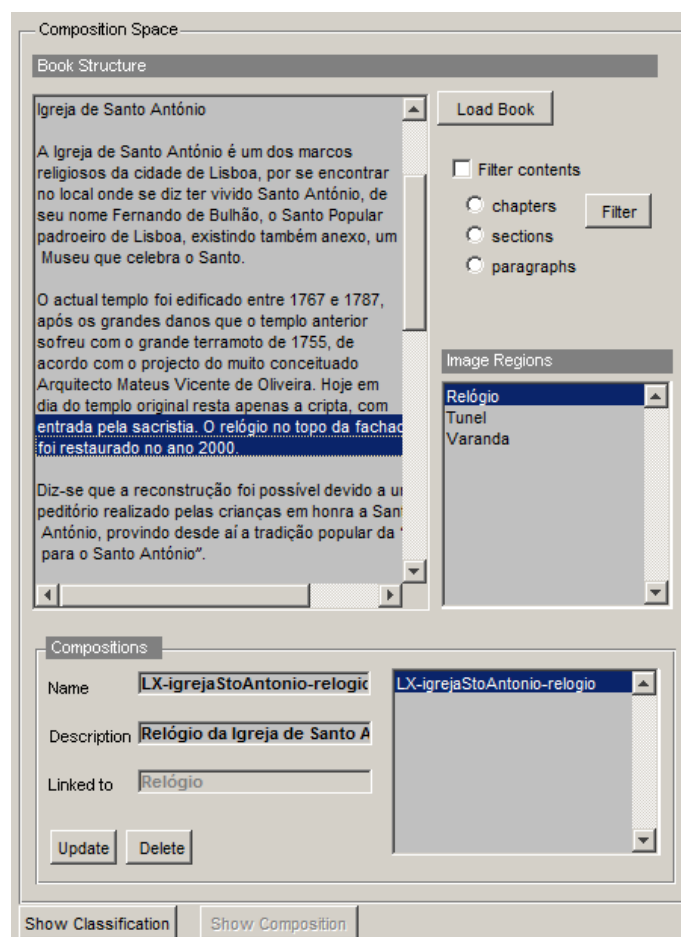


Figura 3.18: Composição de imagem com texto

### 3.4.4 Gestão de conteúdos multimédia

Nesta secção é dada a conhecer a concretização do componente gestor de conteúdos multimédia, relativo à arquitectura da UPI enunciada nesta dissertação. Este componente tem a responsabilidade de garantir a indexação e pesquisa de conteúdos multimédia, de uma forma flexível e eficiente, através de um repositório multimédia construído para o efeito. Pretende-se que este componente indexe as anotações de imagem produzidas pelo componentes de classificação e, as composições de imagem com texto resultantes do componente de composição de imagem. Que garanta ainda, a pesquisa de anotações e composições de imagem realizadas através da ferramenta SICA. De seguida, é descrito com mais detalhe a concretização do repositório multimédia e os componentes de indexação e pesquisa de conteúdos multimédia.

#### *Repositório multimédia*

O repositório multimédia é responsável pelo armazenamento de colecções de imagens e das respectivas anotações e descrições semânticas e sintácticas. Para além disso, apoia o processo de composição, quer armazenando textos de livros, quer indexando as respectivas composições de imagens com texto. Este repositório é concretizado através de uma base de dados *eXist-db*<sup>2</sup>, na qual os elementos multimédia se encontram armazenados em estruturas semelhantes à de um sistema de ficheiros hierárquico.

O *eXist-db*(Meier, 2002) é um sistema de gestão de base de dados *Open Source* inteiramente construído sobre tecnologia XML. A manipulação de conteúdos é feita através de XQuery(Siméon et al., 2007) e XPath(DeRose and Clark, 1999) recorrendo a interfaces *REST*(Fielding, 2000). De modo, a dar suporte à comunicação com a base de dados foram desenvolvidas diversas bibliotecas em JAVA e MATLAB. Estas bibliotecas não são mais, do que a concretização de um *web services* baseado em *REST-Style Web API*<sup>3</sup> que implementa a comunicação entre a ferramenta SICA e a base de dados. Para efeitos de indexação e pesquisa de conteúdos multimédia, na forma de colecções de dados, foram desenvolvidas bibliotecas também em JAVA e MATLAB e, utilizadas as tecnologias XQuery e XPath.

#### *Indexação de conteúdos*

Tanto as anotações de imagens e as próprias imagens, como as composições de imagens com texto e o respectivo texto, são guardadas localmente ou no repositório multimédia. Depois de se carregar uma imagem, de se anotar as regiões pretendidas e de se descrever semanticamente estas anotações, há que guardar a informação resultante destas operações no repositório multimédia. Assim, no que diz respeito ao processo de classificação, o

---

<sup>2</sup><http://exist.sourceforge.net>

<sup>3</sup>[http://www.exist-db.org/exist/devguide\\_rest.xml](http://www.exist-db.org/exist/devguide_rest.xml)

componente de indexação apresentado na Figura 3.19 tem a seu cargo as seguintes responsabilidades:

- a indexação de imagens no repositório: cada imagem é guardada tal como ela é, sem alteração das suas características ou do seu formato, quer seja carregada via Web, quer seja carregada localmente;
- a indexação de imagens anotadas no repositório: por cada imagem, quer seja carregada a partir do repositório, quer seja via Web ou localmente, que for alvo de anotação, é criada uma nova imagem, em tudo, igual à primeira imagem, cujo nome passa a ser um identificador único baseado em *MD5*<sup>4</sup>;
- a indexação de anotações de imagens no repositório: o ficheiro *RDF/XML*<sup>5</sup> de cada imagem anotada é indexado usando, igualmente, um nome único baseado em *MD5*. No caso, de ser carregada na ferramenta, uma imagem que já contenha anotações, este ficheiro é actualizado e indexado com o mesmo nome.



Figura 3.19: Componente de indexação

Este componente de indexação, para além, das responsabilidades acima referidas, tem ainda que lidar com outras, relacionadas com a composição de imagens com texto:

- a indexação de textos no repositório: cada ficheiro de texto é indexado, exactamente tal como está, ora em formato *TXT*, ora em formato *XML* representando o texto de um LFD;
- a indexação de textos compostos com imagens no repositório: se um qualquer ficheiro de texto do repositório for alvo de composição com imagens, é criado um novo ficheiro de texto, em tudo igual, ao primeiro, mas com um nome único baseado em *MD5*;

<sup>4</sup><http://www.ietf.org/rfc/rfc1321.txt>

<sup>5</sup><http://www.w3.org/TR/REC-rdf-syntax>

- a indexação de composições de imagem com texto no repositório: o ficheiro *RDF/XML* com a descrição da composição produzida é indexado usando, igualmente, um nome único baseado em *MD5*.

### ***Pesquisa de conteúdos***

O método de pesquisa de conteúdos no repositório multimédia é feito através de uma chave de pesquisa textual, ou através da escolha do número de resultados de uma pesquisa aleatória. Por um lado, temos uma pesquisa aleatória de imagens, por outro lado, uma pesquisa baseada num texto de pesquisa sobre o nome e sobre os metadados das imagens, sejam estes relacionados com a anotação de imagens ou com a composição de imagens com texto. Tanto num caso, como no outro, os resultados de pesquisa são imagens apresentadas em miniatura e com paginação.

O componente de pesquisa apresentado na Figura 3.20 é realizado através de bibliotecas MATLAB e JAVA e, do uso de *XPath* e *XQuery*, tendo a seu cargo as seguintes responsabilidades:



Figura 3.20: Componente de pesquisa

- a pesquisa de imagens no repositório: a Figura 3.21 apresenta a forma como é concretizada a pesquisa de imagens no repositório. As imagens resultantes da pesquisa no repositório são apresentadas de forma paginada e, a navegação baseia-se num conjunto sequencial de páginas. Cada página inclui, no máximo, seis imagens por página, bastando um simples *click* sobre a imagem pretendida para que esta esteja disponível para análise no espaço de anotação. Dependendo do método de pesquisa, é visualizado um número específico de imagens aleatórias, ou o conjunto de imagens resultante da pesquisa textual baseada no nome da imagem ou nos metadados desta, que resultaram do processo de anotação e classificação de imagens.
- a pesquisa de anotações de imagem no repositório: a concretização da pesquisa de anotação de imagens no repositório é apresentada na Figura 3.22 e é, semelhante ao

que foi descrito no ponto anterior, para o caso de pesquisas simplesmente, de imagens. As imagens resultantes duma pesquisa de anotações de imagem correspondem, apenas, a imagens que já foram alvo de anotação. Para além, da apresentação em miniatura destas imagens, é também apresentado o número de regiões já anotadas nas mesmas. Bastando assim, um *click* sobre a imagem, para se dar continuidade ao processo de anotação de novas regiões de interesse, ou mesmo, remover regiões previamente anotadas, que por algum motivo deixaram de fazer sentido.

- a pesquisa de composições de imagem no repositório: a concretização da pesquisa de composições de imagens com texto no repositório é, em tudo, semelhante, à pesquisa de anotações de imagens. Contudo, no caso da composição, para além de cada miniatura de imagem apresentar o número de anotações que a imagem possui, é apresentado também o número de composições que já foram realizadas nessa imagem. Isto é, cada imagem tem destacado, o número de composições com texto que foi alvo.

A ferramenta SICA funciona de forma independente da bancada de LFDE. No en-

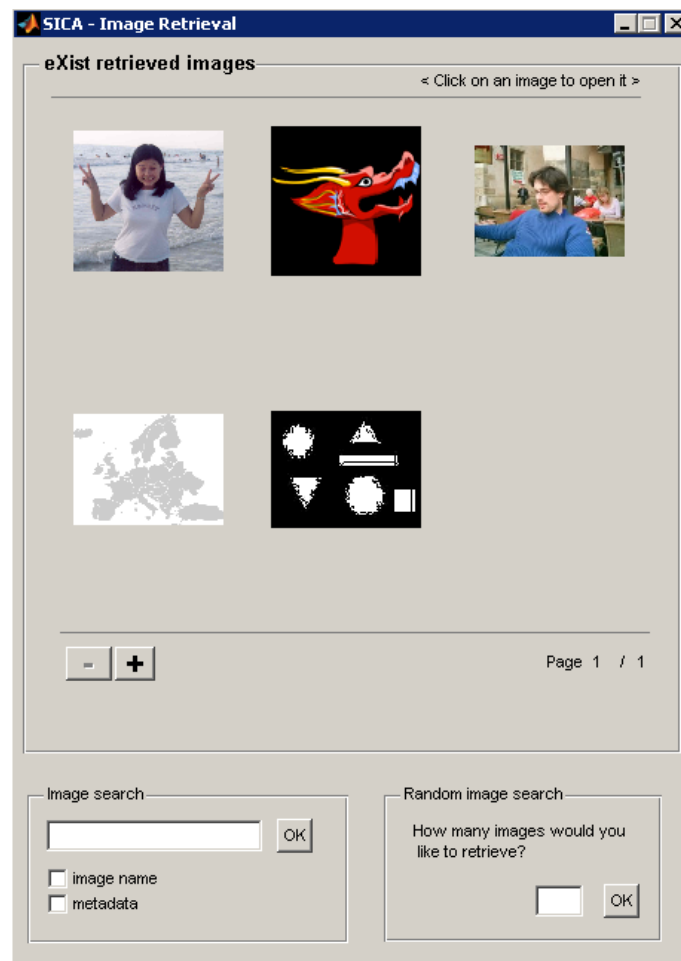


Figura 3.21: Pesquisa de imagens



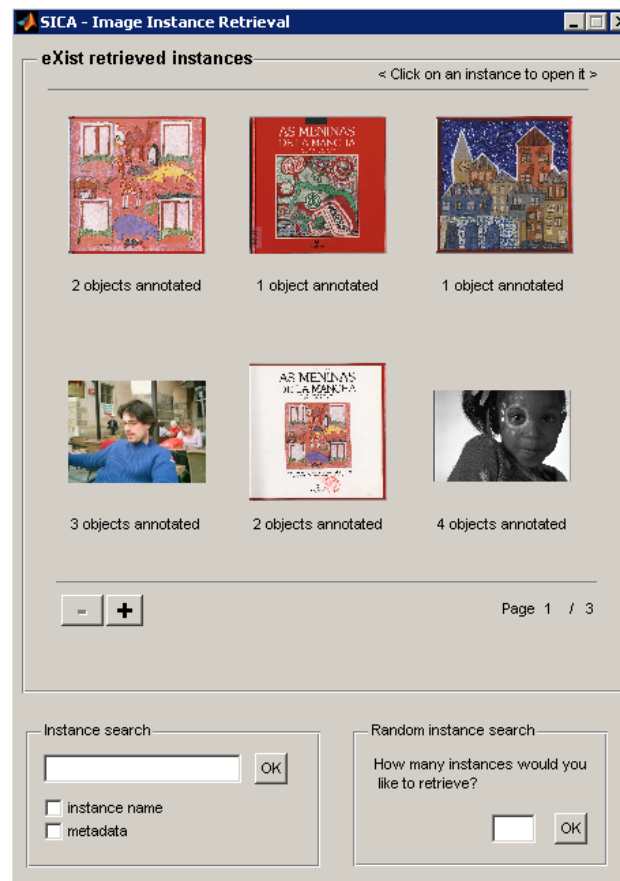


Figura 3.22: Pesquisa de anotação de imagens

tanto, com a criação do repositório multimédia descrito acima, passa a ser possível, a integração dos conteúdos resultantes da sua utilização com a bancada de produção de LFDE. Esta integração assenta, fundamentalmente, na articulação da referida bancada com o repositório multimédia, permitindo referenciar os conteúdos presentes no repositório, na construção de LFDE.

### 3.5 Sumário

Neste capítulo foi descrita a arquitectura genérica de processamento de conteúdos multimédia capaz de solucionar os desafios inerentes à classificação e composição de elementos multimédia. De seguida, foram expostos os casos de uso mais importantes para fazer face a um protótipo que fosse de encontro a estes desafios, mas direccionado para conteúdos baseados em imagens. Como esta dissertação tem por objectivo a construção de uma ferramenta de anotação, classificação e composição de imagens, foi também apresentado o desenho do protótipo aqui proposto e, descrita em porme



# Capítulo 4

## Caso de Estudo

Neste capítulo são apresentados os resultados da aplicação da ferramenta SICA, no processo de construção de um livro falado digital enriquecido específico. Estes resultados resultam da articulação desta ferramenta com a bancada de produção de LFDE e com a ferramenta de leitura *Rich Book Player*(RBP) existentes no projecto RICOBA. O caso de estudo pretende demonstrar as capacidades da ferramenta SICA na construção de LFD e no seu enriquecimento com imagens, assim como, os resultados da reprodução de LFDE no RBP.

Em primeiro lugar são enunciados os requisitos e os procedimentos necessários para construir o LFDE, a partir do do livro original - *As Meninas de La Mancha*. De seguida, são apresentados os mecanismos de preparação necessários para se começar a construir o livro e, por fim são demonstrados os resultados obtidos nas seguintes fases do processo:

1. na classificação e anotação semântica de imagens;
2. na gestão de conteúdos no repositório multimédia;
3. na articulação com a bancada de produção de livros falados digitais;
4. na criação de LFDE com imagens resultantes da utilização da SICA;
5. na reprodução do novo LFDE usando o RBP.

Neste capítulo pode ser observado a concretização da criação de um LFDE, em que as imagens passam a constar no livro, que até então, só tinha texto. Assim, passa a haver uma contextualização do texto presente no livro com as imagens anotadas semanticamente, disponíveis no repositório multimédia. Com a utilização do RBP é também possível, ver e ouvir o livro produzido, para além de se poder interagir com este, das mais variadas formas.

## 4.1 Requisitos

Para construir um LFDE a partir do livro *As Meninas de La Mancha*, é necessário começar por reunir os diversos requisitos, que cada um dos componentes presentes no processo de produção de LFDE necessita para processar a informação pretendida. Assim, a primeira fase começa por reunir toda a informação que o livro em papel compreende, e transformar essa informação em formato digital. Uma vez reunida a informação agora digital, é necessário indexá-la num repositório multimédia. De seguida, entra em cena a utilização da ferramenta SICA, isto é, a anotação e classificação de todas as imagens contidas no livro, assim como, as composições que associam o texto e as imagens. E ainda, a indexação da informação resultante das tarefas de processamento da SICA, no repositório multimédia. Por fim, procede-se à criação do LFDE - *As Meninas de La Mancha* e à sua reprodução na ferramenta de leitura RBP.

Os requisitos para a construção do LFDE a partir do livro *As Meninas de La Mancha* podem ser vistos como pertencentes às seguintes categorias:

### 1. conteúdos digitais genéricos

- texto, imagens, áudio, índices e outras informações descritivas do livro, tal como: o título, a autoria, o editor, o ISBN e a descrição das imagens que o compõem;

### 2. conteúdos digitais compostos

- imagens anotadas, sincronização de áudio com texto e sincronização de imagem com texto;

### 3. repositório multimédia

- indexação e pesquisa dos conteúdos digitais genéricos;

### 4. produção e reprodução de LFDE

- 1), 2) e 3) e ainda, mais alguns conteúdos específicos em XML;
- na reprodução são considerados os requisitos da produção.

A figura 4.1 pretende descrever os artefactos que são requeridos para a construção e produção de LFDE, do ponto de vista da ferramenta SICA, e também dos outros conteúdos que necessariamente tem de ser produzidos e armazenados no repositório multimédia.

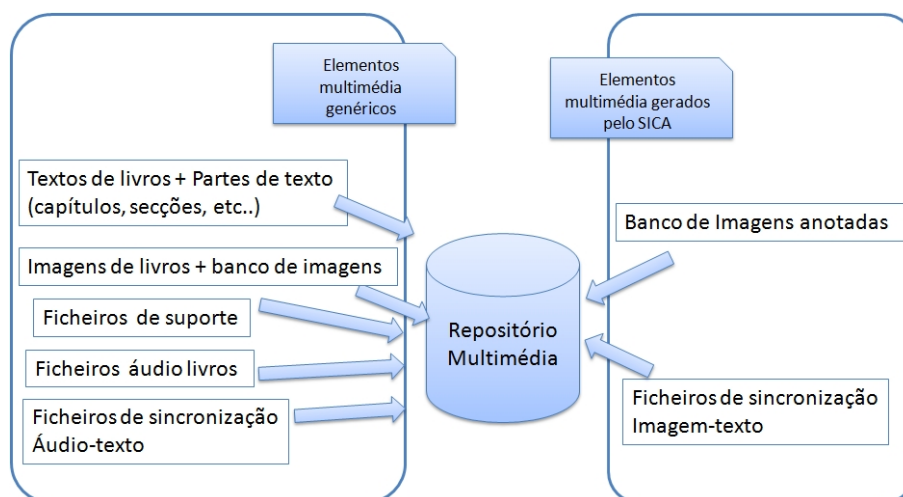


Figura 4.1: Artefactos para a produção de LFDE

## 4.2 Produção e reprodução de LFDE

Nesta secção são apresentados os resultados práticos da aplicação da ferramenta SICA na construção de LFDE com imagens, em articulação com o RBP. Do nosso parceiro de projecto INESC-ID obtiveram-se os seguintes artefactos genéricos em formato digital:

- **texto:** quinze ficheiros de texto em formato TXT que correspondem ao texto da folha de rosto e aos catorze capítulos da obra;
- **imagens:** oitenta e sete ficheiros de imagens em formato JPG resultantes da digitalização completa da obra;
- **áudio:** quinze ficheiros áudio em formato WAV que correspondem à narração do texto da folha de rosto e dos catorze capítulos da obra.

Para além destes conteúdos genéricos, o INESC-ID providenciou ainda, os seguintes artefactos, relativos à composição de áudio com o texto da obra, isto é, à sincronização de áudio com texto. Trata-se de ficheiros de texto em formato LAB que contêm, em cada linha, a indicação do tempo em que cada palavra é falada, tal como é descrito na Figura 4.2. Estes quinze ficheiros estão intimamente ligados a cada um dos ficheiros de texto e de áudio descritos atrás. Resultam da composição dos momentos temporais, em milissegundos em que são narradas cada uma das palavras presentes em cada um dos ficheiros de texto da obra.

Todos os artefactos são armazenados num repositório multimédia, tal como apresentado na Figura 4.3 para, posteriormente, servirem de suporte às fases distintas de criação do LFDE aqui proposto.

Assim, a base de dados XML - *eXists XML Database* é onde todos os conteúdos fornecidos pelo INESC-ID são indexados. A partir do momento em que os conteúdos

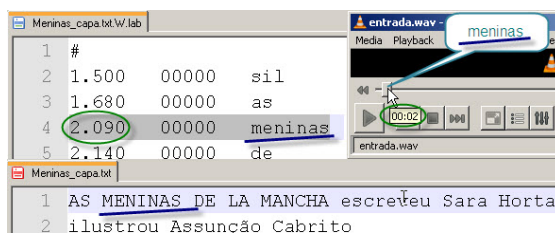


Figura 4.2: Descrição da sincronização de áudio com texto

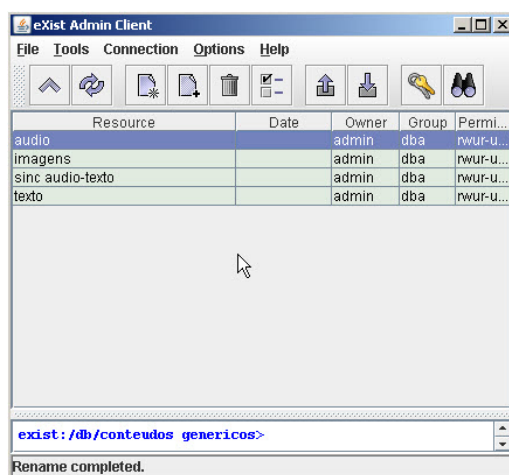


Figura 4.3: Repositório multimédia

genéricos passaram a estar acessíveis no repositório multimédia, tal como é apresentado na Figura 4.4 as tarefas de construção do LFDE podem ser iniciadas.

As tarefas de produção do LFDE começam pela utilização da funcionalidade *Criar livro* disponibilizada pelo RBP, que tem como requisitos de entrada o texto completo da obra e o ficheiro completo de sincronização de áudio com texto. Como resultado do seu processamento, tem-se o ficheiro XML apresentado na Figura 4.5, que combina ambos os conteúdos, permitindo a sincronização do áudio com o texto. Este ficheiro XML apresenta uma estrutura que é a de um livro constituído por parágrafos que contêm frases. Cada frase contém palavras, designadas por unidades. Cada unidade contém uma palavra, o tempo no ficheiro áudio completo, em que a palavra é narrada e o posicionamento da mesma, no texto completo da obra. Para ir de encontro aos requisitos desta funcionalidade é necessário obter os ficheiros completos para cada um dos conteúdos da obra, e outros ficheiros que são necessários para o LFDE ser reproduzido pelo RBP. Deste modo, tem-se que proceder a diversas tarefas de pré-processamento de conteúdos.

Para criar o ficheiro de texto completo - *AsMeninasDeLaMancha.txt*, é usado o comando *copy*, do próprio sistema operativo, de forma a concatenar os diversos ficheiros de texto num único só ficheiro. A Figura 4.6 mostra também um outro ficheiro criado com o mesmo conteúdo deste TXT, mas em formato RTF - *AsMeninasDeLaMancha.rtf*.

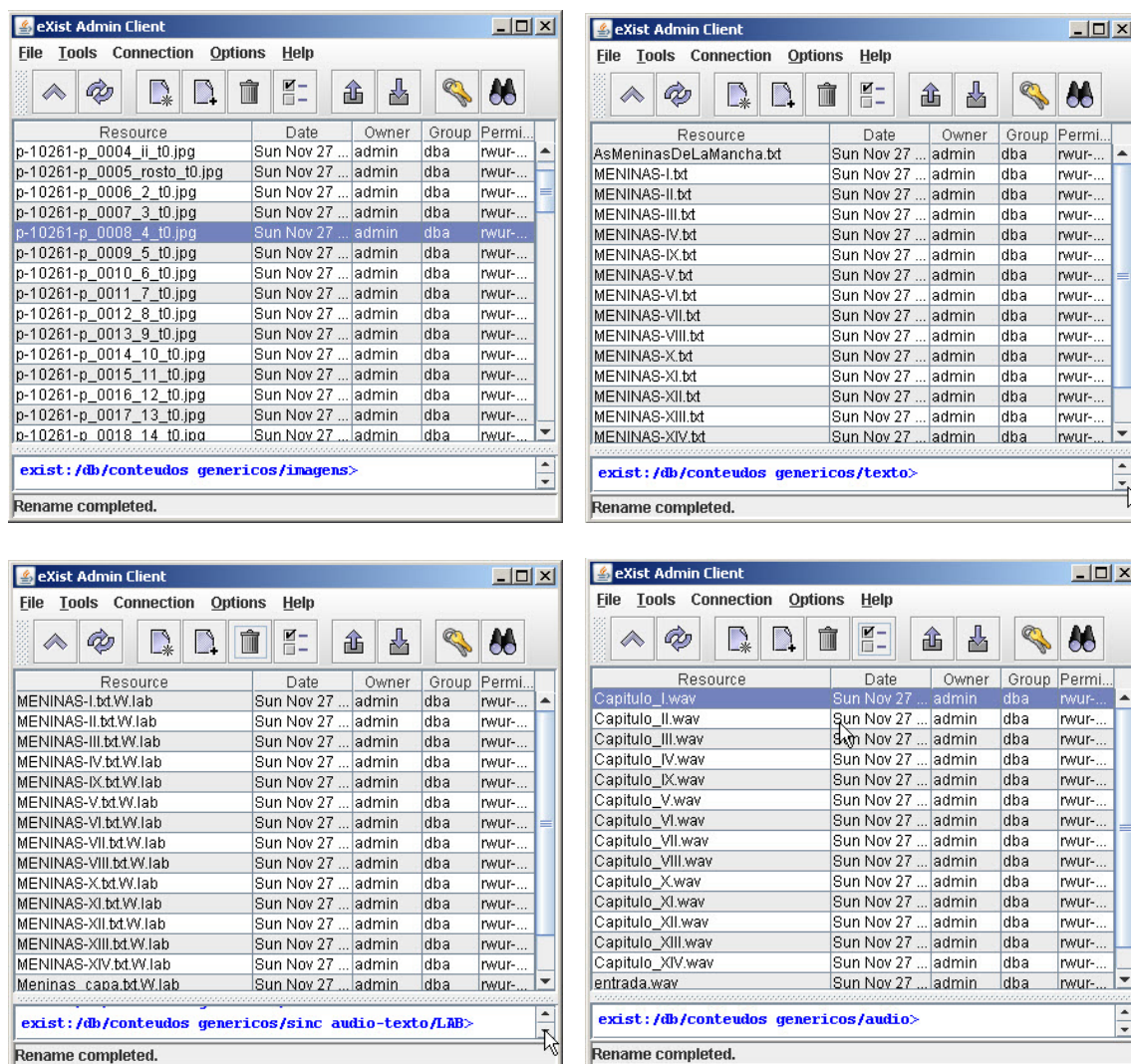


Figura 4.4: Conteúdos no repositório multimédia

- C:\>copy Meninas\_capa.txt + MENINAS-I.txt + MENINAS-II.txt + ... + MENINAS-XIV.txt AsMeninasDeLaMancha.txt

Para ir de encontro à necessidade de se ter um ficheiro de sincronização de áudio com texto completo é usado o *Microsoft Excel*. O conteúdo de cada um dos ficheiros LAB, é copiado para uma folha de *Excel* e os tempos actualizados. Esta actualização de tempos é feita acumulando o tempo da última entrada do ficheiro anterior, à primeira ocorrência do ficheiro seguinte, dando origem ao ficheiro *AsMeninasDeLaMancha.lab*, como pode ser observado na Figura 4.7.

A partir dos artefactos genéricos de áudio é criado o ficheiro de áudio completo. E utilizando a ferramenta *SoX*<sup>1</sup> produz-se um único ficheiro áudio da narração da obra -

<sup>1</sup><http://sox.sourceforge.net>



```

▼<book name="As Meninas de La Mancha" author="Sara Horta Monteiro">
  ▶<paragraph number="1">...</paragraph>
  ▶<paragraph number="2">...</paragraph>
  ▼<paragraph number="3">
    ▶<sentence number="14">...</sentence>
    ▼<sentence number="15">
      ▼<unit number="24">
        <content>mil</content>
        ▼<position>
          <start>1145</start>
          <end>1148</end>
        </position>
        ▼<time>
          <start>16,111</start>
          <end>16,24</end>
        </time>
      </unit>
      ▶<unit number="25">...</unit>
      ▶<unit number="26">...</unit>
      ▶<unit number="27">...</unit>
    </sentence>
  ▶<sentence number="16">...</sentence>
</paragraph>
▶<paragraph number="4">...</paragraph>
▶<paragraph number="5">...</paragraph>
▶<paragraph number="6">...</paragraph>
▶<paragraph number="7">...</paragraph>
▶<paragraph number="8">...</paragraph>
▶<paragraph number="9">...</paragraph>
▶<paragraph number="10">...</paragraph>
▶<paragraph number="11">...</paragraph>
▶<paragraph number="12">...</paragraph>
▶<paragraph number="13">...</paragraph>
▶<paragraph number="14">...</paragraph>
▶<paragraph number="15">...</paragraph>
▶<paragraph number="16">...</paragraph>

```

Figura 4.5: Descrição do XML de sincronização áudio com o texto

*AsMeninasDeLaMancha.wav*, que resulta da concatenação dos diversos ficheiros WAV mencionados atrás.

- C:\>sox ---combine concatenate Entrada.wav Capitulo\_I.wav  
Capitulo\_II.wav ... Capitulo\_XIV.wav AsMeninasDeLaMancha.wav

Para além disto, a tabela de conteúdos do livro é criada consoante as especificidades do RBP. É identificado no áudio completo do livro cada um dos momentos de início de cada um dos capítulos. Tendo em conta, o tipo de nó na árvore do índice e o nome do capítulo, é criado o ficheiro *AsMeninasDeLaMancha.toc* que pode ser visualizado na Figura 4.8.

Depois de se prepararem todos os ficheiros descritos atrás, é necessário tratar-se da criação e configuração de um ficheiro RBK que é, o componente principal que trata de mapear todos os conteúdos que integram o livro, de modo, a que este, passe a ser denominado um LFDE e possa ser reproduzido no RBP. Na Figura 4.9 pode observar-se o conteúdo do ficheiro *AsMeninasDeLaMancha.rbk*.

O ficheiro *Annotations.xml* é o único que não necessita de ser produzido durante as tarefas de pré-processamento descritas anteriormente. Isto deve-se ao facto, de ser um ficheiro meramente indicativo, do local onde deverão ser guardadas todas as anotações que se façam durante a reprodução do livro.



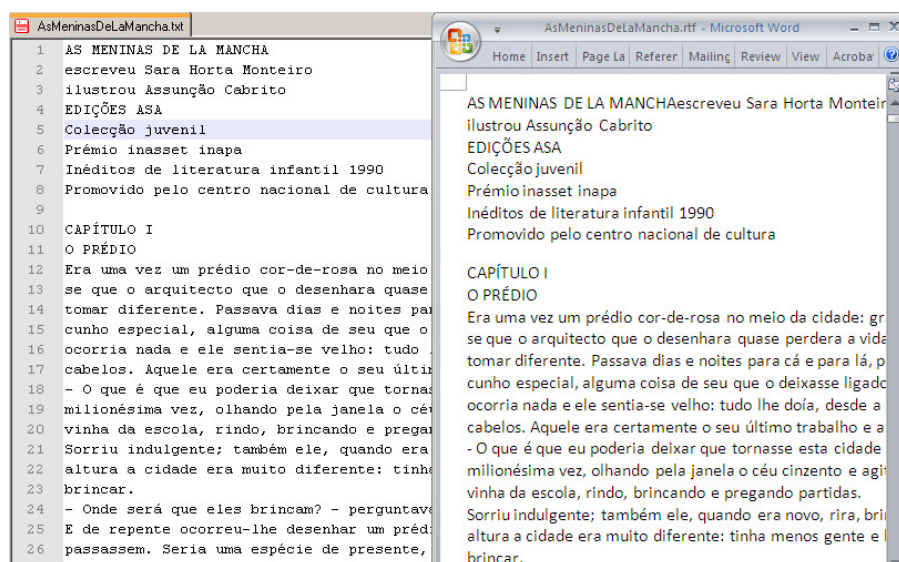


Figura 4.6: Ficheiro TXT e RTF do texto completo

Meninas_capa.txt.W.lab	MENINAS-I.txt.W.lab	AsMeninasDeLaMancha.lab
23 11.010 00000 juvenil	1 0.790 00000 sil	37 18.310 0 promovido
24 12.260 00000 sil	2 1.150 00000 i	38 18.550 0 pelo
25 12.670 00000 prêmio	3 1.770 00000 capítulo	39 18.880 0 centro
26 13.060 00000 inasset	4 2.700 00000 sil	40 19.400 0 nacional
27 14.060 00000 inapa	5 2.820 00000 o	41 19.450 0 de
28 14.410 00000 sil	6 3.380 00000 prédio	42 19.940 0 cultura
29 15.040 00000 inéditos	7 5.250 00000 sil	43 72.230 0 sil
30 15.120 00000 de	8 5.420 00000 era	44
31 15.580 00000 literatura	9 5.580 00000 uma	45 73.390 0 capítulo
32 16.110 00000 infantil	10 5.840 00000 vez	46 74.010 0 i
33 16.240 00000 mil	11 5.940 00000 um	47 74.940 0 sil
34 16.750 00000 novecentos	12 6.250 00000 prédio	48 75.060 0 o
35 16.790 00000 e	13 6.890 00000 cor-de-rosa	49 75.620 0 prédio
36 17.320 00000 noventa	14 6.990 00000 no	50 77.490 0 sil
37 17.780 00000 sil	15 7.210 00000 meio	51 77.660 0 era
38 18.310 00000 promovido	16 7.290 00000 da	52 77.820 0 uma
39 18.550 00000 pelo	17 7.860 00000 cidade	53 78.080 0 vez
40 18.880 00000 centro	18 8.370 00000 sil	54 78.180 0 um
41 19.400 00000 nacional	19 8.900 00000 grande	55 78.490 0 prédio
42 19.450 00000 de	20 9.150 00000 sil	56 79.130 0 cor-de-rosa
43 19.940 00000 cultura	21 9.880 00000 imponente	57 79.230 0 no
44 72.230 00000 sil	22 10.120 00000 sil	58 79.450 0 meio
45 72.240 00000 </s>	23 10.820 00000 magnífico	59 79.530 0 da
46	24 11.740 00000 sil	60 80.100 0 cidade

Figura 4.7: Ficheiro LAB de sincronização de áudio com texto

Até este ponto, fica apenas a faltar, as tarefas relacionadas com a anotação das imagens presentes no livro e, a criação e configuração do ficheiro *imagesList.txt* que faz a sincronização de imagem com o texto. Assim, para anotar as imagens que constam no livro (ver Figura 4.10), usou-se as funcionalidades disponíveis pela ferramenta SICA. Do mesmo modo, que para se criar o ficheiro de sincronização de imagem com texto, se recorreu à composição das imagens anotadas com texto ou partes do texto do livro em causa, usando a mesma ferramenta.

As Figuras 4.11, 4.12, 4.13 e 4.14 demonstram a utilização da ferramenta SICA no processo de anotação de imagens e na composição de imagens com texto.

1	1+FOLHA DE ROSTO+0+0+FOLHA DE ROSTO
2	1+CAPÍTULO I+73,030+0+O PRÉDIO
3	1+CAPÍTULO II+364,650+0+ALICE
4	1+CAPÍTULO III+968,770+0+A DESCOBERTA
5	1+CAPÍTULO IV+1725,480+0+O LABIRINTO
6	1+CAPÍTULO V+2065,410+0+INTERVALO
7	1+CAPÍTULO VI+2269,890+0+A FLORESTA E A
8	1+CAPÍTULO VII+2562,470+0+O PESADELO
9	1+CAPÍTULO VIII+3001,780+0+OS ANIMAIS
10	1+CAPÍTULO IX+3337,700+0+O PRESENTE
11	1+CAPÍTULO X+3538,810+0+O VIZINHO
12	1+CAPÍTULO XI+4365,790+0+O REGRESSO
13	1+CAPÍTULO XII+4547,420+0+MARGARIDA
14	1+CAPÍTULO XIII+5251,420+0+GUSTAVO
15	1+CAPÍTULO XIV+5486,660+0+O PRÉDIO

Figura 4.8: Ficheiro TOC de descrição da tabela de conteúdos

```

1 name=As Meninas de La Mancha
2 author=Sara Horta Monteiro
3 BookFile=AsMeninasDeLaMancha.rtf
4 Sync=File
5 SyncFile=AsMeninasDeLaMancha.lab
6 BookXMLFile=AsMeninasDeLaMancha.xml
7 AudioFile=AsMeninasDeLaMancha.wav
8 ToCFile=AsMeninasDeLaMancha.toc
9 AnnotationsFile=annotations.xml

```

Figura 4.9: Conteúdo do ficheiro *AsMeninasDeLaMancha.RBK*

Figura 4.10: Imagens que constam na obra

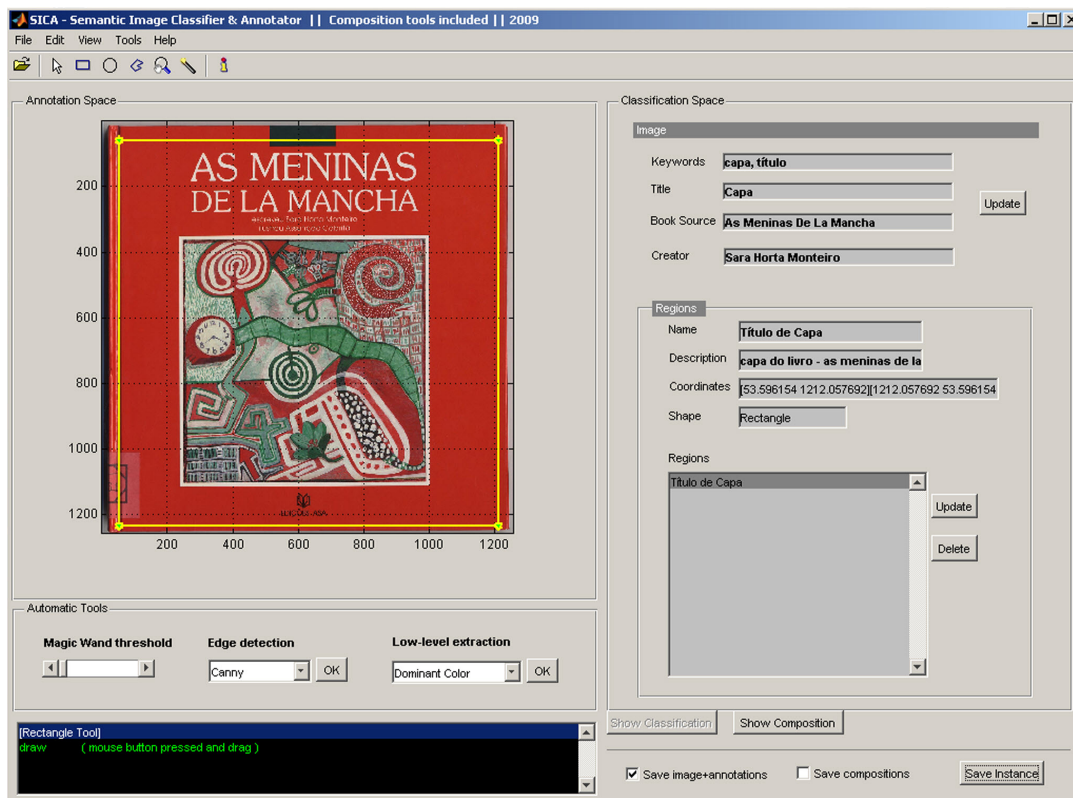


Figura 4.11: Anotação da imagem da Capa do livro

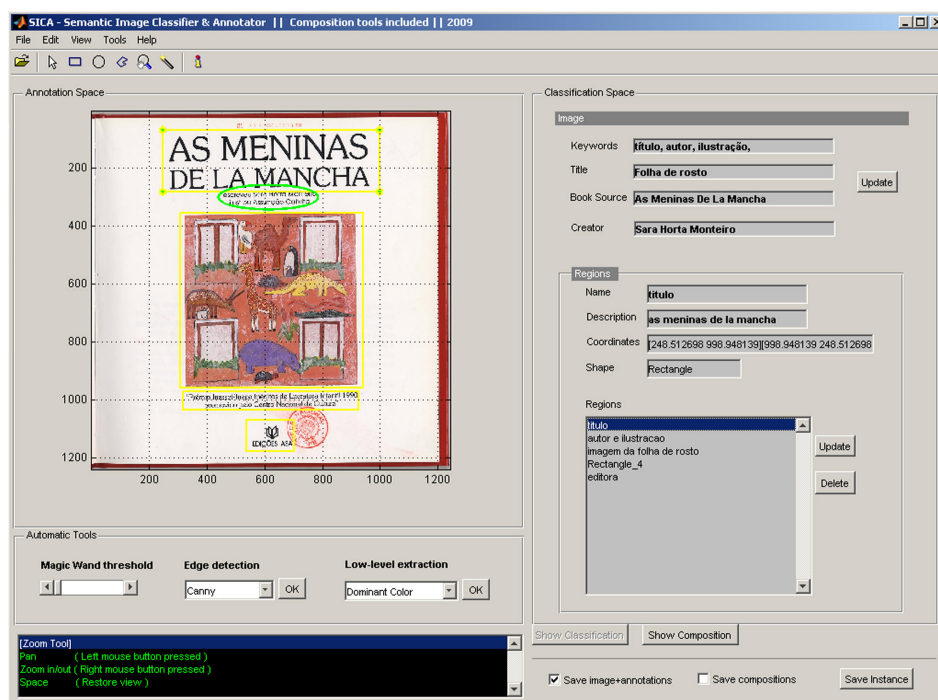


Figura 4.12: Anotação da imagem da Folha de Rosto do livro

O conteúdo do ficheiro *imagesList.txt* apresentado na Figura 4.15, resulta do processamento efectuado pela ferramenta SICA, no que respeita, à sincronização de uma deter-

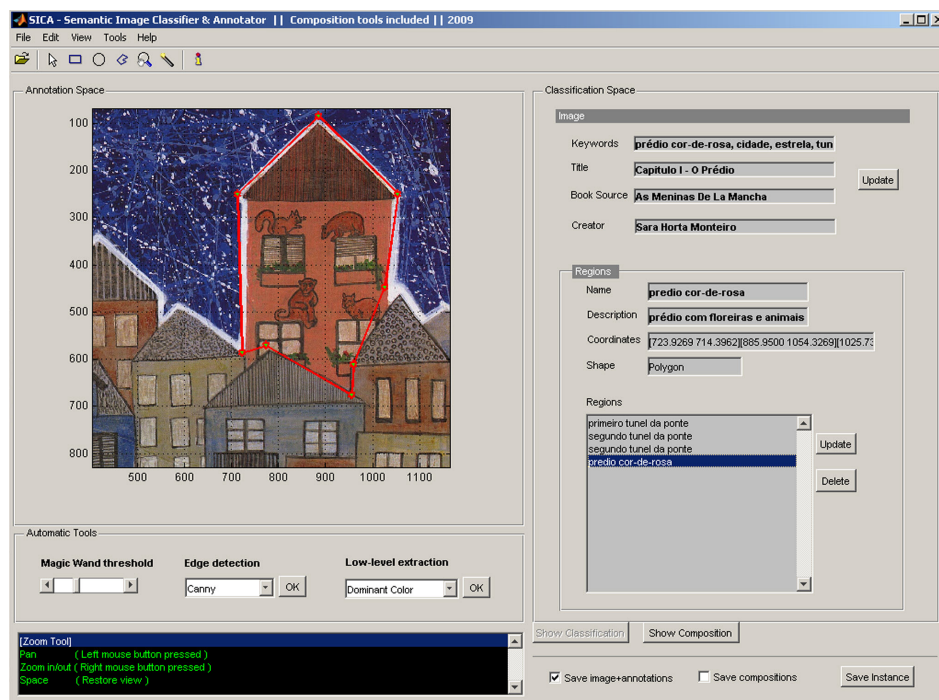


Figura 4.13: Anotação da primeira imagem do Capítulo I

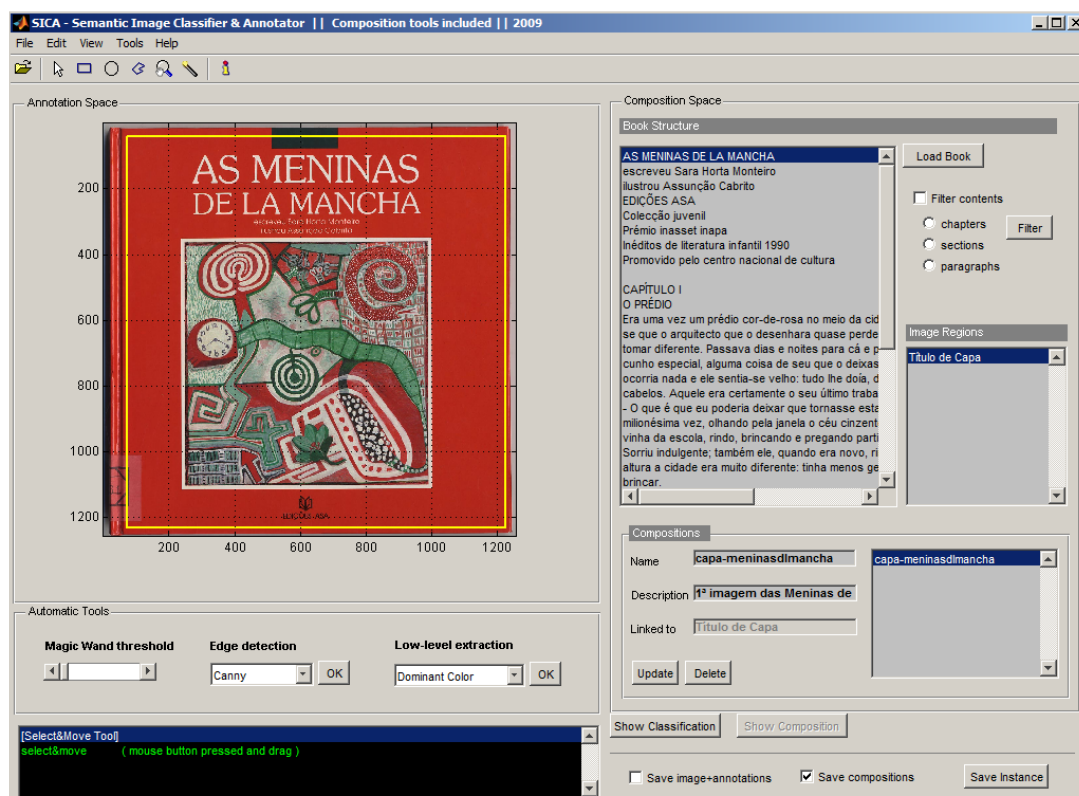


Figura 4.14: Composição da imagem da Capa do livro com o texto

minada imagem em relação ao posicionamento no texto da obra.



imagesList.txt	
1	12
2	p-10261-p_0001_capa_t0.jpg+0+23+Figura 1+Capa
3	p-10261-p_0005_rosto_t0.jpg+24+210+Figura 2+Folha de Rosto
4	p-10261-p_0009_5_t0.jpg+1279+2370+Figura 3+prédios e céu estrelado
5	p-10261-p_0019_15_t0.jpg+3543+3600+Figura 4+caracol peixinhos estrelas mar
6	p-10261-p_0027_23_t0.jpg+4500+4640+Figura 5+instrumentos cor-de-rosa bola relógio
7	p-10261-p_0033_29_t0.jpg+5112+5200+Figura 6+relógio flôr serpente cobra caminhos l
8	p-10261-p_0041_37_t0.jpg+5598+5665+Figura 7+flores bolas cor-de-rosa folhas
9	p-10261-p_0046_42-43_t0.jpg+5900+6100+Figura 8+gato chapéu chupas dinossauro flore
10	p-10261-p_0052_49_t0.jpg+6200+6776+Figura 9+janelas parede prédio animais girafa p
11	p-10261-p_0064_61_t0.jpg+6901+7200+Figura 10+flores cato poltrona chávana torneira p
12	p-10261-p_0074_71_t0.jpg+7328+7400+Figura 11+espiral cavalo cadeira urso folhas
13	p-10261-p_0082_79_t0.jpg+7499+8200+Figura 12+janelas parede prédio animais girafa

Figura 4.15: Ficheiro de sincronização de imagem com texto

Depois de concluídas todas as tarefas de criação do LFDE, o repositório multimédia passa a conter a informação apresentada nas figuras 4.16 e 4.17. Na primeira, é visível a colecção necessária para um livro ser reproduzido pelo RBP, na segunda a colecção de imagens que o vai ser apresentada no livro.

eXist Admin Client				
File Tools Connection Options Help				
Resource				
Date	Owner	Group	Permi...	
Images				
AsMeninasDeLaMancha.lab	admin	dba	rwur-u...	
AsMeninasDeLaMancha.rbk	admin	dba	rwur-u...	
AsMeninasDeLaMancha.rtf	admin	dba	rwur-u...	
AsMeninasDeLaMancha.toc	admin	dba	rwur-u...	
AsMeninasDeLaMancha.txt	admin	dba	rwur-u...	
AsMeninasDeLaMancha.wav	admin	dba	rwur-u...	
AsMeninasDeLaMancha.xml	admin	dba	rwur-u...	
exist: /db/rbplayer/AsMeninasDeLaMancha>				
Rename completed.				

Figura 4.16: Colecção de entrada para o RBP

A reprodução do texto do LFDE - *As Meninas De La Mancha*, agora com imagens, é feita através da ferramenta de leitura RBP como pode ser observado nas Figuras 4.18, 4.19 e 4.20.

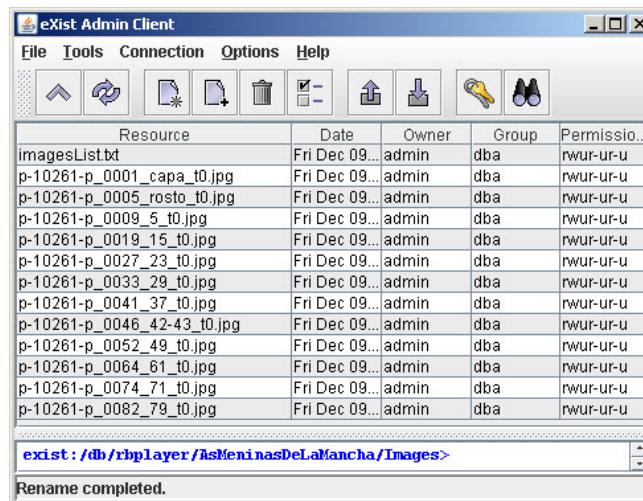


Figura 4.17: Colecção de imagens para o RBP

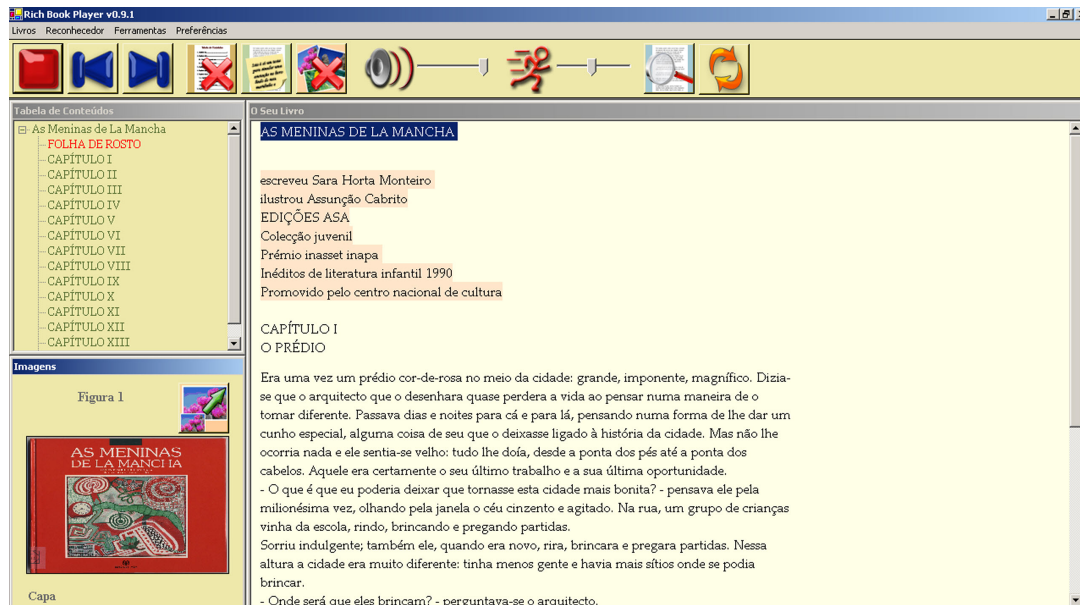


Figura 4.18: RBP a reproduzir o título do LFDE

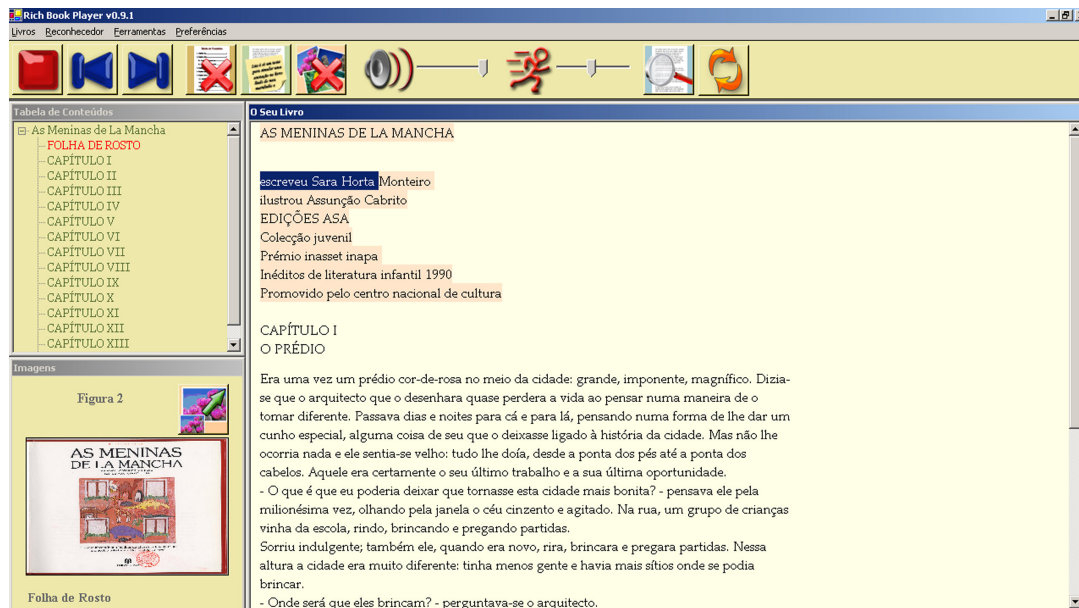


Figura 4.19: RBP a reproduzir a folha de rosto do LFDE

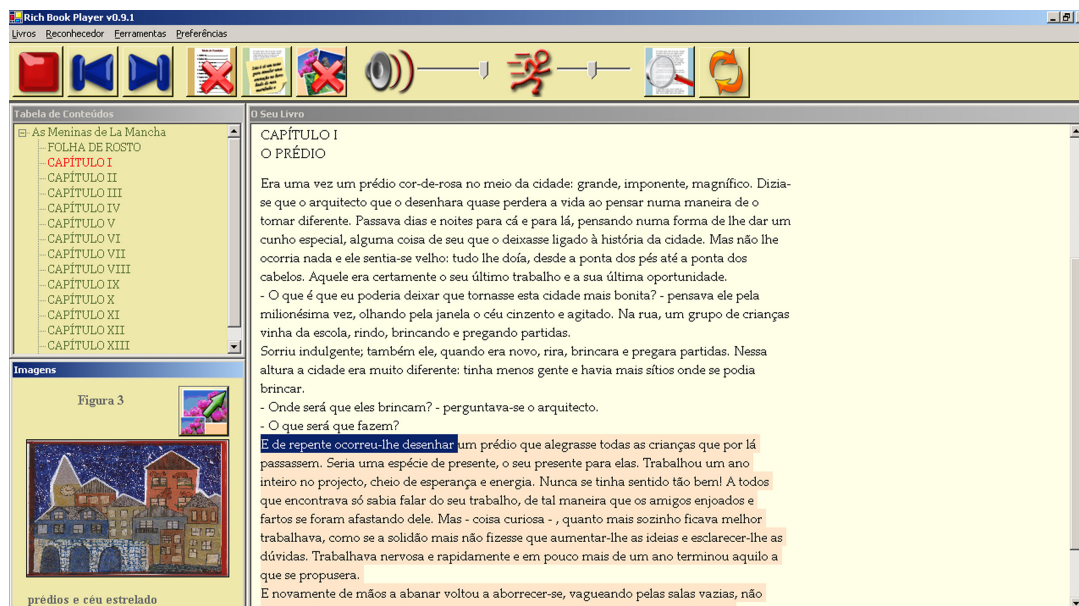


Figura 4.20: RBP a reproduzir o capítulo I do LFDE





# Capítulo 5

## Conclusão

O trabalho desenvolvido cumpriu o plano de trabalhos proposto, com maior ênfase no desenvolvimento da ferramenta de anotação e classificação semântica de imagens permitindo explorar direcções e tendências inicialmente não previstas que incidiram no processo de anotação, classificação e na representação visual dos conteúdos e suas classificações. Para além disso, foram ainda desenvolvidos aspectos relacionados com a composição de imagem com texto e com a integração da ferramenta de classificação com o repositório multimédia. O desenvolvimento da ferramenta SICA teve um impacto fundamental na boa articulação com os restantes componentes desenvolvidos no contexto do projecto Ri-CoBa.

### 5.1 Trabalho Futuro

No contexto da unidade de processamento de imagens, será interessante acrescentar novas funcionalidades ao classificador de imagens. Deve ser analisada a inclusão de um motor de inferências para a ontologia base; um editor de ontologias e um método para assistir o utilizador no processo de extracção de características, de modo, a sugerir nomes para regiões ou objectos.

É ainda desejável fazer evoluir o protótipo de modo a permitir, para além das imagens, a anotação, classificação, e composição de outros tipos de elementos multimédia, como vídeo, áudio, ou até mesmo, o texto. Espera-se que a integração de componentes baseados nestes outros elementos multimédia, no protótipo aqui apresentado, não seja muito complexa, até pelas características próprias do ambiente em que foi desenvolvido, o MATLAB.



# Bibliografia

aceMedia. acemedia homepage, 2006. <http://www.acemedia.org/aceMedia>.

ANSI/NISO. Specifications for the digital talking book, 2002. Available at <http://www.niso.org/standards/resources/Z39-86-2002.html>.

Dave Beckett. RDF/XML syntax specification (Revised). Technical report, W3C, 2 2004. URL <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>.

Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. Scientific American.COM, May 2001. <http://www.starlab.vub.ac.be/teaching/berners-lee.pdf>.

Stephan Bloehdorn, Kosmas Petridis, Carsten Saathoff, Nikos Simou, Vassilis Tzouvaras, Yannis Avrithis, Siegfried Handschuh, Yiannis Kompatsiaris, Steffen Staab, and Michael G. Strintzis. Semantic annotation of images and videos for multimedia analysis. In *Proceedings of the Second European conference on The Semantic Web: research and Applications*, ESWC'05, pages 592–607, Berlin, Heidelberg, 2005. Springer-Verlag. ISBN 3-540-26124-9, 978-3-540-26124-7. doi: 10.1007/11431053\_40. URL [http://dx.doi.org/10.1007/11431053\\_40](http://dx.doi.org/10.1007/11431053_40).

Dan Brickley and Libby Miller. FOAF Vocabulary Specification 0.97. Namespace document, The Friend of a Friend (FOAF) project, January 2010. URL <http://xmlns.com/foaf/spec/20100101.html>.

Jeen Broekstra, Arjohn Kampman, and Frank van Harmelen. Sesame: A generic architecture for storing and querying rdf and rdf schema. In *Proceedings of the First International Semantic Web Conference on The Semantic Web*, ISWC '02, pages 54–68, London, UK, UK, 2002. Springer-Verlag. ISBN 3-540-43760-6. URL <http://dl.acm.org/citation.cfm?id=646996.711426>.

Luís Carriço, Nuno Guimarães, Carlos Duarte, Teresa Chambel, and Hugo Simões. Spoken books: Multimodal interaction and information repurposing. In *Proceedings of*

- HCII'2003, International Conference on Human-Computer Interaction*, pages 680–684, Crete, Greece, 2003.
- Chad Carson, Megan Thomas, Serge Belongie, Joseph M. Hellerstein, and Jitendra Malik. Blobworld: A system for region-based image indexing and retrieval. In *Third International Conference on Visual Information Systems*. Springer, 1999.
- Diamantino Caseiro, Hugo Meinedo, Antonio Serralheiro, Joao Neto, and Isabel Trancoso. Spoken book alignment using wfsts. In *in Proceedings of HLT02. San Diego CA, 2002*. URL <http://www.inesc-id.pt/pt/indicadores/Ficheiros/181.pdf>.
- Ajay Chakravarthy. Cross-media document annotation and enrichment. In *In, Proc. 1 st Semantic Web Authoring and Annotation Workshop (SAAW2006, 2006*.
- Jacob Cybulski and Tanya Linden. Designing multimedia development environments with reuse in mind. In *10th Australasian Conference on Information Systems ACIS'99*, pages 235–246, Wellington, New Zealand, 1999.
- Daisy Consortium. Playback tools, 2006. URL <http://www.daisy.org/tools/playback.asp>. Retrieved January 18, 2006.
- Steven DeRose and James Clark. XML path language (XPath) version 1.0. W3C recommendation, W3C, November 1999. <http://www.w3.org/TR/1999/REC-xpath-19991116>.
- Carlos Duarte and Luis Carrico. Users and usage driven adaptation of digital talking books. In *International Human Computer Interaction Conference, HCII 2005*. Lawrence Erlbaum Associates, Inc., 2005.
- Carlos Duarte and Luis Carrico. A conceptual framework for developing adaptive multimodal applications. In *11th International Conference on Intelligent User Interfaces, IUI'06*. ACM Press., 2006.
- Carlos Duarte, Teresa Chambel, Luís Carriço, Nuno Guimarães, and Hugo Simões. A multimodal interface for digital talking books. In *Proceedings of WWW/INTERNET 2003*, pages 153–160, Algarve, Portugal, November 2003.
- Edd Dumbill. DOAP: Description of a project, 2004.
- Roy Thomas Fielding. Architectural styles and the design of network-based software architectures, 2000.
- David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002. ISBN 0130851981.

- Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing, 2nd Edition*. Prentice Hall, 2002. ISBN 0-201-18075-8.
- Rafael C. Gonzalez, Richard E. Woods, and Steven L. Eddins. *Digital Image Processing Using MATLAB*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003. ISBN 0130085197.
- Guillaume Flandin. XMLTree: an XML toolbox for Matlab, 2003. <http://www.artefact.tk/software/matlab/xml/>.
- Christian Halaschek-Wiener, Jennifer Golbeck, Andrew Schain, Michael Grove, Bijan Parsia, and Jim Hendler. Photostuff - an image annotation tool for the semantic web. In *OPster Proceedings of the 4th International Semantic Web Conference*, 2005. URL [http://www.mindswap.org/~chris/publications/PhotoStuffCR\\_pid83.pdf](http://www.mindswap.org/~chris/publications/PhotoStuffCR_pid83.pdf).
- Siegfried Handschuh. *Creating Ontology-based Metadata by Annotation for the Semantic Web*. Ph.d. thesis (dr. rer. pol.), University of Karlsruhe (TH), Universität Karlsruhe (TH), Institut AIFB, D-76128 Karlsruhe, 2005. Studer, R.; Weinhardt C.
- Siegfried Handschuh and Steffen Staab. Authoring and annotation of web pages in cream. In *Proceedings of the 11th international conference on World Wide Web, WWW '02*, pages 462–473, New York, NY, USA, 2002. ACM. ISBN 1-58113-449-5. doi: 10.1145/511446.511506. URL <http://doi.acm.org/10.1145/511446.511506>.
- Siegfried Handschuh, Steffen Staab, and Fabio Ciravegna. S-cream - semi-automatic creation of metadata. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, EKAW '02*, pages 358–372, London, UK, UK, 2002. Springer-Verlag. ISBN 3-540-44268-5. URL <http://dl.acm.org/citation.cfm?id=645362.650866>.
- Siegfried Handschuh, Steffen Staab, and Alexander Maedche. CREAM - Creating relational metadata with a component-based, ontology-driven annotation framework. In *Proceedings of K-Cap 2001*, pages 76–83, October 2001. URL <http://www.aifb.uni-karlsruhe.de/WBS/sha/papers/kcap2001-annotate-sub.pdf>.
- Till Hanisch. Metadata for electronic documents using the dublin core. In Mehdi Khosrow-Pour, editor, *Encyclopedia of Information Science and Technology (IV)*, pages 1928–1930. Idea Group, 2005. ISBN 1-59140-553-X.
- Lynda Hardman, Dick C. A. Bulterman, and Guido van Rossum. The amsterdam hypermedia model: Adding time, structure and context to hypertext. *Communications of the ACM*, 37:50–62, 1994.

- Jeff Heflin and James Hendler. Searching the web with shoe, 2000.
- Ian Horrocks, Peter F. Patel-Schneider, and Frank Van Harmelen. From shiq and rdf to owl: The making of a web ontology language. *Journal of Web Semantics*, 1:2003, 2003.
- B. Husemann and Gottfried Vossen. Ontology-driven multimedia object management for private users - overview and research issues. *AIS SIGSEMIS Bulletin*, 1(1):29–31, 2004.
- ISO. Iso 19115. In Shashi Shekhar and Hui Xiong, editors, *Encyclopedia of GIS*, page 600. Springer, 2008. ISBN 978-0-387-30858-6.
- Bernd Jahne. *Digital Image Processing, 5th revised and extended edition*. Springer-Verlag, 2002. ISBN 0-471-37407-5.
- Jeff Ayars, Dick Bulterman, *et al.* Synchronized Multimedia Integration Language (SMIL 2.0). W3C Rec., 2001a. <http://www.w3.org/TR/SMIL2>.
- Jeff Ayars, Dick Bulterman, *et al.* Synchronized Multimedia Integration Language (SMIL 2.0). W3C Rec., 2001b. <http://www.w3.org/TR/SMIL2>.
- JEITA. Exchangeable image file format for digital still cameras: Exif version 2.2. JEITA CP-3451, April 2002. URL <http://www.bibsonomy.org/bibtex/2ebd00c16dc2f30f091e4439fffb15ad54/dret>.
- José Kahan, Marja-Riitta Koivunen, Eric Prud'Hommeaux, and Ralph R. Swick. Annotea: An open rdf infrastructure for shared web annotations, 2001.
- Eva Klien. A rule-based strategy for the semantic annotation of geodata. *T. GIS*, 11(3): 437–452, 2007. URL <http://ceur-ws.org/Vol-187/23.pdf>.
- R. Lopes, H. Simoes, C. Duarte, and L. Carrico. Rich Digital Books for the Web. *Web Information Systems and Technologies (WEBIST'2007)*, 2007. URL <http://www.di.fc.ul.pt/~lmc/pubs/pdfs/2007webist-rdbweb-rlopes.pdf>.
- Rui Lopes, Luis Carrico, and Carlos Duarte. DiTaBBu - Automating the production of time-based hypermedia content. In *15th International World Wide Web Conference 2006 (accepted for publishing as poster paper)*, 2006.
- Wolfgang Meier. exist: An open source native xml database. In *Web, Web-Services, and Database Systems*, pages 169–183, 2002.
- IEEE MultiMedia. Mpeg-7: The generic multimedia content description standard, part 1. *IEEE MultiMedia*, 9(2):78–87, April 2002. ISSN 1070-986X. doi: 10.1109/93.998074. URL <http://dx.doi.org/10.1109/93.998074>.

- Wayne Niblack, Ron Barber, William Equitz, Myron Flickner, Eduardo H. Glasman, Dragutin Petkovic, Peter Yanker, Christos Faloutsos, and Gabriel Taubin. The qbic project: Querying images by content, using color, texture, and shape. In *Storage and Retrieval for Image and Video Databases (SPIE)*, pages 173–187, 1993.
- Patrick Schmitz and Jin Yu and Peter Santangeli. Timed Interactive Multimedia Extensions for HTML (HTML+TIME). W3C Note, 1998. <http://www.w3.org/TR/NOTE-HTMLplusTIME>.
- A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: content-based manipulation of image databases. *Int. J. Comput. Vision*, 18(3):233–254, 1996. ISSN 0920-5691.
- Kosmas Petridis, Dionysios Anastasopoulos, Carsten Saathoff, Norman Timmermann, Yiannis Kompatsiaris, and Steffen Staab. M-ontomat-annotizer: image annotation linking ontologies and multimedia low-level features. In *Proceedings of the 10th international conference on Knowledge-Based Intelligent Information and Engineering Systems - Volume Part III, KES'06*, pages 633–640, Berlin, Heidelberg, 2006. Springer-Verlag. ISBN 3-540-46542-1, 978-3-540-46542-3. doi: 10.1007/11893011\_80. URL [http://dx.doi.org/10.1007/11893011\\_80](http://dx.doi.org/10.1007/11893011_80).
- POSC/EIA/61042/04. Projecto RiCoBa - Livros com Conteudos Ricos para Todos (POSC/EIA/61042/04), 2004. URL <http://hcm.di.fc.ul.pt/ricoba/>.
- POSI/PLP/34252/99. Projecto IPSOM - Indexacao, Integracao e Pesquisa de Som em Documentos Multimedia (POSI/PLP/34252/99), 1999. URL <http://hcm.di.fc.ul.pt/ipsom/>.
- William K. Pratt. *Digital Image Processing: PIKS Inside, 3th Edition*. John Wiley & Sons, Inc., 2001. ISBN 0-471-37407-5.
- Eric Prud'hommeaux and Andy Seaborne. Sparql query language for rdf (working draft). Technical report, W3C, March 2007. URL <http://www.w3.org/TR/2007/WD-rdf-sparql-query-20070326/>.
- Miguel Rodrigues, Luis Carrico, Carlos Duarte, Nuno Guimaraes, Antonio Serralheiro, and Isabel Trancoso. New ways to read digital talking books. In *International Conference Applied Computing 2006, IADIS*, 2006.
- Patrick Schmitz, Jin Yu, and Peter Santangeli. Timed Interactive Multimedia Extensions for HTML (HTML+TIME). W3C Note, 1998. <http://www.w3.org/TR/NOTE-HTMLplusTIME>.
- Linda G. Shapiro and George C. Stockman. *Computer Vision*. Prentice Hall, 2001. ISBN 0-13-030796-3.

- Thomas Sikora. The MPEG-7 visual standard for content description - an overview. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):696–702, June 2001.
- Jérôme Siméon, Don Chamberlin, Daniela Florescu, Scott Boag, Mary F. Fernández, and Jonathan Robie. XQuery 1.0: An XML query language. W3C recommendation, W3C, January 2007. <http://www.w3.org/TR/2007/REC-xquery-20070123/>.
- N. Simou, V. Tzouvaras, Y. Avrithis, G. Stamou, and S. Kollias. A Visual Descriptor Ontology for Multimedia Reasoning. In *In Proceedings of Workshop on Image Analysis for Multimedia Interactive Services*. WIAMIS 05, April 13-15 2005. URL <http://www.image.ece.ntua.gr/php/savepaper.php?id=342>.
- C. M. Sperberg-McQueen, François Yergeau, Eve Maler, Jean Paoli, and Tim Bray. Extensible markup language (XML) 1.0 (fifth edition). W3C proposed edited recommendation, W3C, February 2008. <http://www.w3.org/TR/xml>.
- S. Staab, A. Mädche, and S. Handschuh. *An Annotation Framework for the Semantic Web*. FZI report. Inst. AIFB, Univ., 2001. URL <http://books.google.pt/books?id=cfTecQAACAAJ>.
- Stanford Medical Informatics. The protégé ontology editor and knowledge acquisition system, 2006. <http://protege.stanford.edu/>.
- Stanford University. Ontolingua homepage, 2006. <http://www.ksl.stanford.edu/software/ontolingua/>.
- V. S. Subrahmanian. *Principles of multimedia database systems*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1998. ISBN 1-55860-466-9.
- Mischa M. Tuffield, Stephen Harris, David P. Dupplaw, Ajay Chakravarthy, Christopher Brewster, Nicholas Gibbins, Kieron O'Hara, Fabio Ciravegna, Derek Sleeman, Nigel R. Shadbolt, , and Yorick Wilks. Image annotation with photocopain. In *Proceedings of the First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, Edinburgh, May 2006. URL <http://image.ntua.gr/swamm2006/resources/paper09.pdf>. held as part of 15th World Wide Web Conference (22-26 May, 2006).
- Jacco van Ossenbruggen, Joost Geurts, Frank Cornelissen, Lloyd Rutledge, and Lynda Hardman. Towards second and third generation web-based multimedia. *The Tenth International World Wide Web Conference in Hong Kong*, pages 479–488, May 1-5 2001.
- VRA. The VRA core categories, version 4.0. Visual Resources Association, 2002. URL <http://www.vraweb.org/projects/vracore4>.



- W3C OWL Guide. OWL Web Ontology Language: Overview, 2004. <http://www.w3.org/TR/owl-guide/>.
- W3C RDF Concepts. Resource Description Framework (RDF): Concepts and Abstract Syntax, 2004. <http://www.w3.org/RDF/>.
- W3C Semantic Web. W3C Semantic Web, 2007. <http://www.w3.org/2001/sw/>.
- W3C Semantic Web Activity. SEMANTICWEB.ORG. Semantic Web Community Portal, 2008. <http://www.w3.org/2001/sw/>.
- W3C SWBPD MM Task Force. Multimedia Annotation on the Semantic Web (MM), 2006. <http://www.w3.org/2001/sw/BestPractices/MM/>.
- David Wood. Kowari: A platform for semantic web storage and analysis. In *In XTech 2005 Conference*, pages 05–0402, 2005.

